Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Desarrollo de un motor de ejecución de restricciones de expresiones de SNOMED CT

Trabajo Fin de Grado

Grado en Ingeniería Informática

Autor: Vicente Miguel Giménez Solano

Tutores: Montserrat Robles Viejo y José Alberto Maldonado Segura

[2016-2017]

Agradecimientos

Mi más sincero agradecimiento a Montserrat Robles, directora del Grupo de Informática Biomédica (IBIME) del instituto ITACA de la Universitat Politècnica de València, por darme la oportunidad de trabajar en su grupo y por su inestimable ayuda, en muchos sentidos.

A José Alberto Maldonado, por transmitirme una pequeña parte de sus profundos conocimientos en informática médica, por sus buenas ideas y por guiarme.

A David Moner, Diego Boscá, Santiago Salas, Christian Ponce, Alejandro Mañas, Estíbaliz Parcero y al resto de investigadores del grupo IBIME, por su apoyo, conocimientos, profesionalidad, cercanía e inmejorable trato.

Resumen

En el área de los sistemas de información sanitarios, dado que la información se encuentra repartida formando islas independientes, es primordial construir sistemas interoperables que sean capaces de transmitir información entre ellos. Un punto clave es conseguir un alto grado de interoperabilidad semántica, gracias a la cual los sistemas entienden la información que les es transmitida y son capaces de trabajar con ella. En este sentido, la principal debilidad actual es la falta de coordinación entre los modelos de información clínicos y los modelos terminológicos para definir el significado y el contenido de los datos clínicos. La organización SNOMED International, consciente de esta problemática, ha desarrollado recientemente el Lenguaje de Restricciones de Expresiones de SNOMED CT. Gracias a este lenguaje, es posible definir subconjuntos de conceptos clínicos que servirán para definir enlaces terminológicos de contenido entre los modelos de información clínicos y terminologías médicas. En este trabajo se describe una implementación de un motor de ejecución para dicho lenguaje, cuyo objetivo final es el enlace terminológico avanzado entre arquetipos y SNOMED CT, como pilar fundamental para conseguir sistemas semánticamente interoperables.

Palabras clave: interoperabilidad semántica, arquetipo, enlace terminológico, enlace semántico, enlace de contenido, enlace de valor, subconjuntos de SNOMED CT.

Abstract

In the clinical information systems area it is primordial to build interoperable systems able to transmit information between them. It is necessary because clinical information is stored and divided in separate islands. A crucial issue is to achieve high levels of semantic interoperability for transmitting and understanding information between systems. Nowadays, one of the weaknesses when working in this direction is the lack of a coordinated use of information models and terminological models to define the meaning and content of clinical data. SNOMED International organisation is aware of this problem and has recently developed the SNOMED CT Expression Constraint Language to specify subsets of concepts. These subsets are used in content terminological binding between clinical information models and terminological models. In this work we describe an implementation of an execution engine for this language. Our final objective is to allow advanced terminological binding between archetypes and SNOMED CT as a fundamental pillar to get semantically interoperable systems.

Keywords : semantic interoperability, archetype, terminological binding, semantic binding, content binding, value binding, SNOMED CT subsets.

Tabla de contenidos

1. Justificación 17
2. Objeto y objetivos 19
3. Introducción 21
4. Interoperabilidad semántica 23
5. Representación de la información clínica 25
   5.1 Historia Clínica Electrónica 25
   5.2 Estándares de arquitecturas de HCE 27
   5.3 Arquitecturas de HCE: el modelo dual 30
6. Terminología SNOMED CT 35
7. Enlace terminológico 49
8. Materiales y métodos 51
   8.1 Lenguaje de Restricciones de Expresiones de SNOMED CT 51
   8.2 Almacenamiento de la base de datos de SNOMED CT 68
9. Resultados 71
   9.1 Implementación de un módulo Java de importación de SNOMED CT 71
      9.1.1 Tiempos medios de creación de la base de datos 72
   9.2 Desarrollo de un motor de ejecución de restricciones de expresiones de SNOMED CT 73
      9.2.1 Generalidades 73
      9.2.2 Validación de las restricciones de expresiones 77
      9.2.3 Traducción al lenguaje Cypher de Neo4j 78
      9.2.4 Fases de ejecución 78
      9.2.5 Optimizaciones 79
      9.2.6 Interfaz de usuario 80
      9.2.7 Visualización de los grafos resultantes 81
      9.2.8 Otras funcionalidades 83
      9.2.9 Tiempos medios de ejecución 84
10. Conclusiones y trabajo futuro 87
11. Referencias 89

Definiciones

- **Interoperabilidad técnica:** capacidad de dos o más sistemas de comunicarse a nivel físico (cableado, clavijas, protocolos...). Se ocupa de la transferencia de bytes.

- **Interoperabilidad sintáctica:** capacidad de dos o más sistemas de transferirse datos y documentos. Garantiza que la estructura de los documentos sea la correcta.

- **Interoperabilidad organizativa:** tiene como base las reglas de negocio. Trata de la cooperación posible entre distintas organizaciones bajo un mismo contexto y unos mismos flujos de trabajo.

- **Interoperabilidad semántica:** capacidad de dos o más sistemas de intercambiar información de manera que puedan entenderla y trabajar con ella como si fuera suya propia.

- **Historia clínica electrónica (HCE o EHR):** conjunto de documentos electrónicos donde se registra el historial clínico de un paciente a lo largo de su vida.

- **Modelo de información clínico:** estructura de datos estándar para el almacenamiento y comunicación de la HCE, como openEHR, ISO 13606 o HL7 CDA.

- **Modelo de información clínico detallado:** estructura de datos construida a partir de un modelo de referencia estandarizado cuyo propósito principal es la comunicación de los datos de la HCE. Por ejemplo, arquetipos, plantillas o formularios.

- **Arquetipo:** se ocupa de representar conceptos clínicos de una manera formal. Se enlazan con terminología clínica. Son una combinación jerarquizada de estructuras del modelo de referencia restringida para ajustarse a la definición del concepto modelado.

- **Instancia de un arquetipo:** datos reales de un paciente introducidos siguiendo la estructura, las restricciones y los tipos de datos que marca un arquetipo determinado y cuyo propósito es poder ser almacenado y comunicado de una manera formal y estandarizada.

- **Modelo dual (o de dos niveles):** modelo de información clínica que separa la parte que no varía con el tiempo, es decir, la información, de la parte que sí evoluciona, es decir, el conocimiento. Por una parte está el modelo de referencia y por otra el modelo clínico detallado, como los arquetipos.

- **CEN EN ISO 13606:** es una norma del Comité Europeo de Normalización (CEN) y también ha sido aprobada como norma ISO. Su propósito es lograr la interoperabilidad semántica en la comunicación de la HCE.

- **openEHR:** es un estándar abierto que se ocupa del almacenamiento y comunicación de la información clínica mediante informes de HCE. openEHR, además, es una comunidad virtual que trabaja en aras de conseguir la interoperabilidad universal.

- **HL7 CDA:** HL7 Clinical Document Architecture (CDA) es un estándar basado en XML cuyo propósito es especificar la codificación, estructura y semántica de documentos clínicos para su intercambio.

- **LOINC:** Logical Observation Identifiers Names and Codes (LOINC) es una base de datos y estándar universal para identificar pruebas de laboratorio.

- **CIE (o ICD):** la Clasificación Internacional de Enfermedades (CIE) es una clasificación de enfermedades, síntomas, hallazgos anormales, etc. En la actualidad va por su décima versión CIE-10. Es el sistema de codificación de enfermedades más utilizado en los sistemas de HCE en el sistema de salud español.

- **Modelo terminológico o terminología médica/clínica:** es un conjunto de términos que representan los conceptos en el campo específico de la medicina.

- **Ontología médica:** es una descripción formal de los conceptos y relaciones en el dominio de la biomedicina, donde se especifican los significados y las relaciones jerárquicas entre los términos y conceptos de dicho domino.

- **SNOMED CT:** Systematized Nomenclature of Medicine – Clinical Terms es la terminología clínica integral, multilingüe y codificada de mayor amplitud, precisión e importancia desarrollada en el mundo. Se estructura como un grafo dirigido y acíclico. Es por tanto una ontología médica.

- **Modelo de conceptos de SNOMED CT:** conjunto de reglas que rigen el modo en que los conceptos de SNOMED CT pueden relacionarse mediante sus atributos. Tanto las expresiones post-coordinadas como las restricciones de expresiones deben definirse siguiendo sus reglas para ser semánticamente correctas. Se compone de dominios, atributos y rangos.

- **Modelo de conceptos de SNOMED CT procesable por ordenador (MRCM):** el Machine Readable Concept Model es una versión del modelo de conceptos de SNOMED CT procesable por ordenador. Está expresado mediante el lenguaje de restricciones de expresiones de SNOMED CT.

- **Concepto pre-coordinado:** es un concepto definido por sus relaciones de atributo y jerárquicas, es decir, por su definición lógica. Los conceptos pre-coordinados forman parte de la terminología y no es necesario crearlos.

- **Concepto post-coordinado:** a diferencia de los pre-coordinados, los conceptos post-coordinados son conceptos que no existen en la terminología. Se crean atendiendo a los requisitos particulares siguiendo las reglas definidas en el modelo conceptual para que sean semánticamente correctos.

- **Expresiones de SNOMED CT:** existen dos tipos de expresiones, las pre-coordinadas, formadas por un solo identificador, y las post-coordinadas, formadas por más de un identificador. A nivel sintáctico, las expresiones post-coordinadas se construyen con base en la gramática composicional de SNOMED CT y, a nivel semántico, siguiendo el modelo conceptual.

- **Restricciones de Expresiones de SNOMED CT:** son reglas computables cuyo propósito es definir un subconjunto de conceptos clínicos. Para ello se sirve de operadores de restricción para navegar por las jerarquías de SNOMED CT, refinamientos para filtrar por atributos y operadores lógicos, entre otros.

- **Lenguaje de Restricciones de Expresiones de SNOMED CT:** es el lenguaje utilizado para definir las restricciones de expresiones de SNOMED CT.

- **Subconjunto (de conceptos clínicos):** conjunto de conceptos clínicos que, por lo general, pertenecen al mismo subdominio. Por ejemplo, enfermedades pulmonares asociadas a edemas, tipos de diabetes mellitus, hallazgos clínicos, procedimientos radiológicos en el corazón, etc.

- **Subconjunto extensional:** subconjunto definido mediante una lista de conceptos. Por ejemplo, diabetes tipo I, diabetes tipo II, etc.

- **Subconjunto intensional (o por comprensión):** subconjunto definido mediante una expresión. Por ejemplo, 'todos los tipos de diabetes'.

- **Enlace terminológico:** asociación entre un elemento del modelo de información clínico, como por ejemplo, un arquetipo, y un concepto o conjunto de conceptos de una terminología clínica, como SNOMED CT. Su principal función es definir el significado y el contenido del modelo de información de manera inequívoca en aras de conseguir un alto grado de interoperabilidad semántica.

- **Enlace semántico:** es uno de los dos tipos de enlace terminológico en el que se asocia un elemento del arquetipo con un concepto de la terminología para definir el significado del elemento de manera inequívoca.

- **Enlace de contenido (o de valor):** es otro tipo de enlace terminológico donde se asocia un elemento del arquetipo con un conjunto de conceptos terminológicos para definir su posible contenido. Es decir, el elemento del arquetipo solo podrá contener como valor, un concepto de dicho conjunto. De lo contrario, el arquetipo será semánticamente inconsistente.

- **Sección ontológica (o terminológica) de un arquetipo:** es la sección del Archetype Definition Language (ADL) destinada a la definición tanto en lenguaje natural como formal (i.e. procesable por ordenador) de los enlaces terminológicos, tanto semánticos como de contenido.

Índice de figuras

Número	Descripción	Página
Figura 1	Ejemplo de dos sistemas donde cada uno “entiende” un significado distinto para la palabra “planta”	24
Figura 2	Los tres pilares sobre los que se asienta la interoperabilidad semántica	24
Figura 3	Vista general de la HCE y sus componentes (fuente: Johns Hopkins University)	26
Figura 4	Diagrama del modelo dual	30
Figura 5	Esquema general del modelo dual	32
Figura 6	Analogía Lego para modelo de referencia y arquetipo	33
Figura 7	Analogía Lego para modelo de referencia y arquetipos	33
Figura 8	Ejemplo de arquetipo ISO 13606: Análisis de sangre	33
Figura 9	Distintas extensiones locales de SNOMED CT	35
Figura 10	Mapa con los distintos países donde se usa de SNOMED CT	35
Figura 11	Fragmento de SNOMED CT con conceptos, relaciones y descripciones	36
Figura 12	Ejemplo de concepto y sus descripciones asociadas, tanto la completa como los sinónimos (el preferido y los aceptables)	37
Figura 13	Ejemplo de dos conceptos y sus subtipos (nótese la polijerarquía)	37
Figura 14	Definición lógica del concepto 425548001 |absceso de corazón|	38
Figura 15	Distintos tipos de grafos	38
Figura 16	Algunas de las jerarquías principales de SNOMED CT	39
Figura 17	Este es el aspecto que presenta SNOMED CT, donde los puntos negros son los conceptos y las líneas grises son las relaciones de subtipo. Obsérvese que las jerarquías principales corresponden a las zonas más oscuras	39
Figura 18	Vista general del modelo lógico de SNOMED CT	40
Figura 19	Definición lógica del concepto 2704003 |enfermedad aguda|	40
Figura 20	Definición lógica del concepto 64572001 |enfermedad|	41
Figura 21	Definición lógica del concepto 69449002 |acción de un fármaco|	41
Figura 22	Partes del modelo conceptual de SNOMED CT	41
Figura 23	Ejemplos de relaciones correctas e incorrectas según el modelo conceptual	42
Figura 24	Ejemplos de relaciones correctas según el modelo conceptual	42
Figura 25	Definición lógica del concepto 174041007 |apendicectomía laparoscópica de emergencia|	43
Figura 26	Modelo lógico de la gramática composicional de SNOMED CT	45
Figura 27	Resumen del modelo lógico de SNOMED CT	47
Figura 28	Resumen del modelo conceptual, expresiones y restricciones de expresiones de SNOMED CT	48
Figura 29	Ejemplo de enlace semántico entre un formulario con datos del paciente y los procedimientos quirúrgicos aplicados, y conceptos de SNOMED CT	49

Figura 30	Ejemplo de enlace de contenido entre un formulario con datos del paciente y los procedimientos quirúrgicos aplicados, y un subconjunto extensional de SNOMED CT	50
Figura 31	Ejemplo de enlace de contenido entre un formulario con datos del paciente y los procedimientos quirúrgicos aplicados, y un subconjunto intensional de SNOMED CT	50
Figura 32	Comparación entre la gramática composicional y el lenguaje de restricciones	51
Figura 33	Modelo abstracto de una restricción de expresiones de SNOMED CT	54
Figura 34	Componentes de una restricción de expresiones	54
Figura 35	Modelo lógico del lenguaje de restricciones de expresiones	55
Figura 36	Conjunto de hallazgos clínicos	62
Figura 37	Conjunto de hallazgos clínicos unido al conjunto de procedimientos	63
Figura 38	Conjunto de tipos de diabetes mellitus con sitio del hallazgo la glándula pituitaria	63
Figura 39	Conjunto de hallazgos clínicos causados por algún tipo de organismo	64
Figura 40	Conjunto de hallazgos clínicos causados por algún tipo de organismo y cuyo proceso patológico es un tipo de proceso infeccioso	64
Figura 41	Conjunto de hallazgos clínicos causados exactamente por tres tipos de organismos	65
Figura 42	Conjunto de todas las sustancias que son agentes causales de hallazgos clínicos	65
Figura 43	Conjunto de hallazgos clínicos que no están causados por ningún tipo de organismo	66
Figura 44	Ejemplo de un subconjunto de nodos de SNOMED CT con sus propiedades y sus relaciones jerárquicas (incluyendo el cierre transitivo) y de atributo. Cada nodo contiene su código, su descripción completa, su profundidad (mínima) en el grafo, su número de descendientes y la jerarquía a la que pertenece. Cada atributo contiene su código, su descripción completa y el número de grupo	69
Figura 45	Interfaz del motor de ejecución de restricciones de expresiones de SNOMED CT tras ejecutar < 19829001 |disorder of lung| AND < 301867009 |edema of trunk|	77
Figura 46	Visualización gráfica de << 73211009 | diabetes mellitus |, sin utilizar ForceAtlas2 (izquierda) y utilizándolo (centro). Una vista detallada parcial de la sub-jerarquía diabetes mellitus tipo I se muestra en el lado derecho de la figura	81
Figura 47	Representación gráfica para el subconjunto de las enfermedades pulmonares asociadas a edema	82
Figura 48	Representación en árbol de n niveles del subconjunto definido por la restricción de expresiones < 19829001 | enfermedad pulmonar | : 116676008 |morfología asociada| = <<79654002 | edema |	83

Índice de tablas

Número	Descripción	Página
Tabla 1	Resumen de los estándares de intercambio de datos y mensajería	28
Tabla 2	Resumen de algunos estándares de terminología	28
Tabla 3	Operadores y funciones requeridas en el lenguaje de restricciones	53
Tabla 4	Operadores de restricción del lenguaje de restricciones de expresiones de SNOMED CT	62
Tabla 5	Operadores lógicos del lenguaje de restricciones de expresiones de SNOMED CT	62
Tabla 6	Pasos en la creación de la base de datos de SNOMED CT y tiempos medios en segundos	73
Tabla 7	Funciones soportadas y no soportadas en el motor de ejecución del lenguaje de restricciones de expresiones de SNOMED CT	76
Tabla 8	Las seis restricciones de expresiones de SNOMED CT que se han ejecutado en el motor de ejecución para el cálculo de los tiempos medios de ejecución	85
Tabla 9	Tiempos medios de ejecución de las seis consultas (en milisegundos)	85

1. Justificación

La interoperabilidad semántica, según el Comité Europeo de Normalización (CEN), es el estado que existe entre dos aplicaciones cuando, en relación a una tarea específica, una de las aplicaciones es capaz de recibir datos de la otra y realizar satisfactoriamente dicha tarea, y todo ello sin que intervenga un operador externo. Dicho de otra manera, para que dos sistemas sean semánticamente interoperables, la información ha de fluir entre ellos de manera que el significado original se mantenga inalterado. Cada uno de estos sistemas ha de entender lo que recibe del otro y reaccionar acorde automáticamente. En el contexto sanitario, la interoperabilidad semántica es un requisito necesario para que la información dispersa de un paciente, generada en distintos lugares, distintos sistemas y en distintos instantes temporales, pueda ser compartida y esté a disposición de los profesionales clínicos siempre que sea necesaria e incluso pueda ser utilizada a efectos estadísticos o de investigación. No obstante, como se puede advertir, no es sencillo alcanzar un nivel alto de interoperabilidad semántica. Existen tres pilares fundamentales sobre los que se asienta: por una parte, los modelos de información clínicos, consistentes en estándares de almacenamiento y comunicación de la historia clínica electrónica (i.e. modelos de referencia HL7 CDA, EN ISO 13606 y openEHR) y los modelos de información clínicos detallados (i.e. arquetipos, plantillas, interfaces visuales, etc.). Y, por otra parte, los modelos terminológicos (i.e. terminologías médicas como SNOMED CT, LOINC o CIE). Además, es imprescindible llevar a cabo un enlace terminológico entre los modelos de información y las terminologías. Existen dos tipos de enlace terminológico: el enlace semántico, que proporciona significado inequívoco a las estructuras de información contenidas en el modelo de información mediante un enlace entre un elemento del modelo de información y un término pre o post-coordinado de la terminología (por ejemplo, el elemento *diabetes mellitus* de un arquetipo corresponde al concepto *73211009 | diabetes mellitus (trastorno)* | de SNOMED CT); y el enlace de contenido (o enlace de valor), que restringe el conjunto de posibles valores codificados de la terminología, susceptibles de ser asociados a un elemento del modelo de información (por ejemplo, el elemento *procedimiento quirúrgico* de un arquetipo está asociado al subconjunto de los procedimientos quirúrgicos de SNOMED CT susceptibles de ser seleccionados en dicho elemento, como *110468005 | cirugía ambulatoria (procedimiento) |*, *711364006 | cirugía con asistencia robótica (procedimiento) |*, *56306000 | cirugía estética (procedimiento) |*, y hasta un total de unos 20.000 procedimientos quirúrgicos). Para lograr el enlace semántico basta con definir el enlace dentro del propio modelo de información clínico detallado entre el elemento en cuestión y el concepto terminológico. Si bien en la literatura científica se han abordado estrategias para crear enlaces semánticos de manera manual o semiautomática, el enlace de contenido no ha sido estudiado en profundidad. Para llevar a cabo el enlace de contenido se requiere un mecanismo de definición de subconjuntos de conceptos terminológicos, como el recientemente creado Lenguaje de Restricciones de Expresiones de SNOMED CT por parte de la organización de desarrollo de estándares terminológicos SNOMED International.

En este sentido, es necesario implementar un motor de ejecución del lenguaje de restricciones de SNOMED CT para poder definir subconjuntos de términos médicos que

se asociarán, dentro del arquetipo, a estructuras de datos contenidas en él. Gracias a esta asociación y gracias a los enlaces semánticos, el significado y contenido del arquetipo se mantendrá inalterado sea cual sea el sistema que haga uso del mismo, lo cual está alineado y es un requisito necesario para lograr sistemas de información clínicos semánticamente interoperables, con todas las ventajas que ello conlleva, como la mejora en la calidad de la continuidad asistencial y la disminución del riesgo de estar consultando información parcial del paciente.

2. Objeto y objetivos

El objeto del presente Trabajo Final de Grado es la obtención del título de Graduado en Ingeniería Informática, expedido por la Universitat Politècnica de València.

El motor de ejecución de restricciones de expresiones de SNOMED CT está implementado en Java (se ha utilizado la plataforma Eclipse) y está disponible vía interfaz web en http://diebosto2.pc.upv.es:8888/SnomedQuery/ para su uso libre. La interfaz ha sido desarrollada con el framework para aplicaciones web open-source para el desarrollo de aplicaciones web, Vaadin. El presente trabajo se centra únicamente en la implementación del motor. La parte correspondiente al desarrollo de la interfaz (así como un servicio web disponible para realizar llamadas al motor) ha sido llevada a cabo con el apoyo de miembros del Grupo de Informática Biomédica (IBIME) del instituto ITACA de la Universitat Politècnica de València, donde ha sido llevado a cabo este trabajo.

El desarrollo del motor de ejecución tiene tres objetivos principales:

- El primer objetivo es **soportar la definición de enlaces terminológicos** entre modelos de información clínicos detallados, tales como arquetipos, y terminologías clínicas, en concreto SNOMED CT. En particular, gracias a la implementación de un motor de ejecución de restricciones de expresiones de SNOMED CT, es posible especificar subconjuntos de términos de SNOMED CT que servirán para realizar el enlace de contenido entre el arquetipo y la terminología SNOMED CT. Enlazar el arquetipo con terminología clínica, tanto mediante enlace semántico como de contenido, es definir inequívocamente el significado del modelo de información y restringir el conjunto de posibles términos asociados a un elemento del arquetipo, respectivamente. Establecer estos enlaces es ir en la dirección de construir sistemas de información clínicos semánticamente interoperables.

- El segundo objetivo es la **validación de instancias de arquetipos que contienen enlaces de contenido** en su sección ontológica. Gracias al motor de ejecución, es posible evaluar a verdadero o falso si un término forma parte de un determinado subconjunto. Cuando se define un enlace de contenido sobre el valor de un atributo en un arquetipo, el valor de dicho atributo ha de estar incluido, para que los datos de la instancia sean consistentes, en un determinado conjunto de valores. Por ejemplo, se puede especificar que el valor de un atributo ha de pertenecer al subconjunto de las infecciones del tracto respiratorio. De este modo, si el atributo hace referencia, por ejemplo, a una gripe, los datos serán consistentes y la instancia será válida, dado que gripe forma parte del subconjunto especificado. Pero si el atributo hace referencia, por ejemplo, a una fractura de fémur, los datos serán inconsistentes, dada la incompatibilidad de este concepto con el subconjunto de las infecciones del tracto respiratorio.

- El tercer objetivo se pone de manifiesto a la hora de especificar reglas o restricciones que afectan a dos o más entidades del arquetipo (i.e. afirmaciones a

nivel global del arquetipo). Estas restricciones se definen en la sección de reglas del arquetipo y algunas de ellas involucran el resultado de consultar un contexto de datos determinado, como puede ser una terminología clínica y en concreto SNOMED CT. Por tanto, gracias al motor de ejecución es posible **dar soporte a restricciones a nivel global del arquetipo** que servirán para enriquecer el arquetipo con conocimiento del dominio para ser capaces de medir su nivel de consistencia, entendida como una dimensión de la calidad de una instancia. En resumen, con el desarrollo de este motor de ejecución se persigue, en última instancia, aumentar el nivel de calidad de las historias clínicas electrónicas.

3. Introducción

El presente Trabajo Final de Grado tiene como propósito el desarrollo de un motor de ejecución que da soporte al Lenguaje de Restricciones de Expresiones de SNOMED CT. Gracias a este lenguaje, es posible definir de manera intensional (i.e. por comprensión) subconjuntos de conceptos terminológicos que, a la postre, podrán ser usados en el enlace terminológico de contenido (también llamados de valor) entre los modelos de información clínicos, tales como arquetipos, y la ontología médica SNOMED CT. Gracias a la utilización del modelo dual como modelo estándar para el almacenamiento y comunicación de la Historia Clínica Electrónica (HCE), y al enlace terminológico, es posible conseguir un alto grado en interoperabilidad semántica, que se define como la capacidad que tienen los sistemas de transmitir información entre ellos de modo que puedan usarla como suya propia, es decir, comprendiendo su significado de manera automática.

El contenido del presente trabajo empieza, tras abordar la justificación y los objetivos, en el punto 4, donde se explican los diferentes tipos de interoperabilidad entre sistemas y se hace hincapié en la interoperabilidad semántica.

En los sucesivos puntos se acometen los pormenores de los elementos necesarios para conseguir un nivel alto en interoperabilidad semántica. Se desgranan por puntos y se profundiza lo necesario en ellos. Concretamente, el punto 5.1 aborda los aspectos generales de la HCE para, a continuación, dar paso al punto 5.2, sobre estándares de arquitecturas de HCE. Seguidamente, en el punto 5.3, se trata un modelo particular para almacenar y comunicar el contenido de la HCE: el modelo dual.

Siguiendo con los pilares donde se asienta la interoperabilidad semántica llegamos al punto 6, donde se explican los fundamentos de la ontología médica SNOMED CT.

El punto 7 está dedicado a los dos tipos de enlace terminológico: enlace semántico y enlace de contenido.

El punto 8.1 aborda el Lenguaje de Restricciones de Expresiones de SNOMED CT para definir subconjuntos terminológicos de manera intensional, como lenguaje al que se da soporte en el motor de ejecución, cuyo desarrollo centra el objetivo de este trabajo. Asimismo, en el punto 8.2 se trata la metodología seguida para almacenar el sustrato de SNOMED CT en una base de datos orientadas a grafos de Neo4j.

Los puntos 9.1 y 9.1.1 se centran en la implementación de un módulo en lenguaje Java para importar de manera automática las distintas ediciones de SNOMED CT y en un análisis de los tiempos medios de creación de la base de datos, respectivamente.

Finalmente, el punto 9.2 aborda de manera detallada el desarrollo del motor de ejecución para definir restricciones de expresiones de SNOMED CT que, una vez evaluadas (i.e. ejecutadas) retornan el listado relativo al subconjunto de conceptos clínicos definidos en la restricción. Se hace hincapié en todos los aspectos relevantes seguidos en el transcurso de la implementación del motor, tales como la validación sintáctica y semántica en el punto 9.2.2; la traducción de la restricción en lenguaje de restricciones de expresiones

de SNOMED CT al lenguaje de consultas sobre base de datos de grafos de Neo4j, Cypher, en el punto 9.2.3; las distintas fases de ejecución de las restricciones en el punto 9.2.4; las optimizaciones en los dominios de las restricciones siguiendo el modelo de conceptos de SNOMED CT en el punto 9.2.5; a continuación se dan unas pinceladas de la interfaz gráfica del motor de ejecución en el punto 9.2.6; el punto 9.2.7 está dedicado a la visualización gráfica de los subconjuntos obtenidos en forma de grafos, burbujas y árboles; el punto 9.2.8 está dedicado a diversas funcionalidades presentes en la interfaz gráfica del motor, tales como la elección de idioma, de sustrato de SNOMED CT, el tipo de sintaxis (i.e. corta o larga), la conexión de los resultados con el navegador de SNOMED International, etc.; y, finalmente, en el punto 9.2.9 se lleva a cabo una evaluación de los tiempos de respuesta del motor de ejecución a partir de una serie de restricciones de expresiones que son ejecutadas un determinado número de veces en aras de calcular los tiempos medios de ejecución del motor.

El trabajo finaliza con un apartado dedicado a las conclusiones y al trabajo que se prevé realizar en el futuro, punto 10, y con el punto 11, que alberga las referencias y el material bibliográfico empleado en la realización de este Trabajo Final de Grado.

4. Interoperabilidad semántica

Existen cuatro tipos de interoperabilidad:

- La *interoperabilidad técnica* trata la conexión entre los distintos sistemas. Se ocupa de establecer los enlaces físicos y lógicos para que los sistemas tengan la capacidad de transmitir información entre ellos: cables, clavijas, protocolos de comunicación, redes inalámbricas. Normas como 802.3, 802.11, TCP/IP, HTTP, la especificación Bluetooth, Zigbee, los niveles bajos de la familia ISO 11073, SOAP, etc., son las que se utilizan para conseguir la interoperabilidad técnica. En definitiva, permite la transferencia de bytes.

- La *interoperabilidad sintáctica* se encarga de hacer posible la transferencia de documentos. Se ocupa de que la estructura de los documentos o los mensajes intercambiados sea la correcta. No obstante, no ejecuta ninguna comprobación de que la información intercambiada tenga sentido o un significado concreto. Por tanto, solo interviene en el formato de los ficheros que se intercambian o de los tipos de datos que se utilizan. Algunas normas empleadas son: XML, las especificaciones para tipos de datos como la TS 14796 de CEN o la ISO 21090, las especificaciones de mensajes de las versiones 2.x de HL7 o los modelos de referencia de HL7 V3 o de UNE-EN ISO 13606, aunque estos últimos son también la base de la interoperabilidad semántica, como veremos después.

- La *interoperabilidad organizativa* tiene como base las reglas de negocio. Trata de la cooperación posible entre distintas organizaciones bajo un mismo contexto y unos mismos flujos de trabajo. Se están empezando a dar los primeros pasos en lo que refiere a interoperabilidad organizativa, sobre todo en el área de la normalización: la norma EN 12967 HISA (Health Informatics Service Architecture) en cuya primera parte trata el punto de vista de la empresa, y sobre todo la norma UNE-EN ISO 13940 (sistema de conceptos para dar soporte a la continuidad asistencial).

- La *interoperabilidad semántica* se define como el estado que existe entre dos aplicaciones cuando, en relación a una tarea específica, una de las aplicaciones es capaz de recibir datos de la otra y realizar satisfactoriamente dicha tarea, y todo ello sin que intervenga un operador externo. O, en otras palabras, para que dos sistemas sean semánticamente interoperables, la información ha de fluir entre ellos de manera que el significado original se mantenga inalterado. Cada uno de estos sistemas ha de entender lo que recibe del otro y reaccionar acorde automáticamente. En este contexto, conviene subrayar que en los sistemas de información clínicos, la información no está concentrada, sino que se halla formando islas independientes. Es por ello que el riesgo de consultar información parcial de un paciente se hace patente, con todos los riesgos que ello puede conllevar.

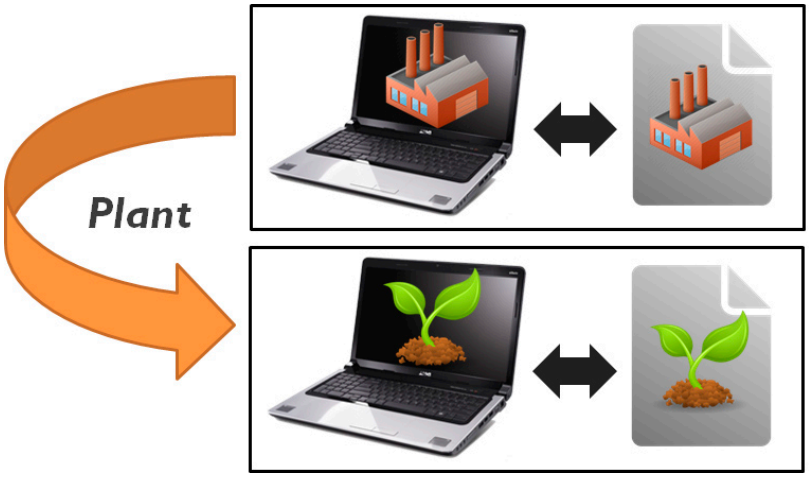

Figura 1. Ejemplo de dos sistemas donde cada uno "entiende" un significado distinto para la palabra "planta"

En el contexto sanitario, la interoperabilidad semántica es un requisito necesario para que la información dispersa de un paciente, generada en distintos lugares, distintos sistemas y en distintos instantes temporales, pueda ser compartida y esté a disposición de los profesionales clínicos siempre que sea necesaria e incluso pueda ser utilizada a efectos estadísticos o de investigación. En cualquier caso, alcanzar un nivel alto de interoperabilidad semántica es una tarea compleja.

Existen tres pilares fundamentales para lograr sistemas semánticamente interoperables: por una parte, los modelos de información clínicos, consistentes en estándares de almacenamiento y comunicación de la historia clínica electrónica (i.e. modelos de referencia HL7 CDA, EN ISO 13606 y openEHR) y los modelos de información clínicos detallados (i.e. arquetipos, plantillas, interfaces visuales, etc.). Y, por otra parte, los modelos terminológicos (i.e. terminologías médicas como SNOMED CT, LOINC o CIE). Además, es imprescindible llevar a cabo un enlace terminológico entre los modelos de información y las terminologías [1].

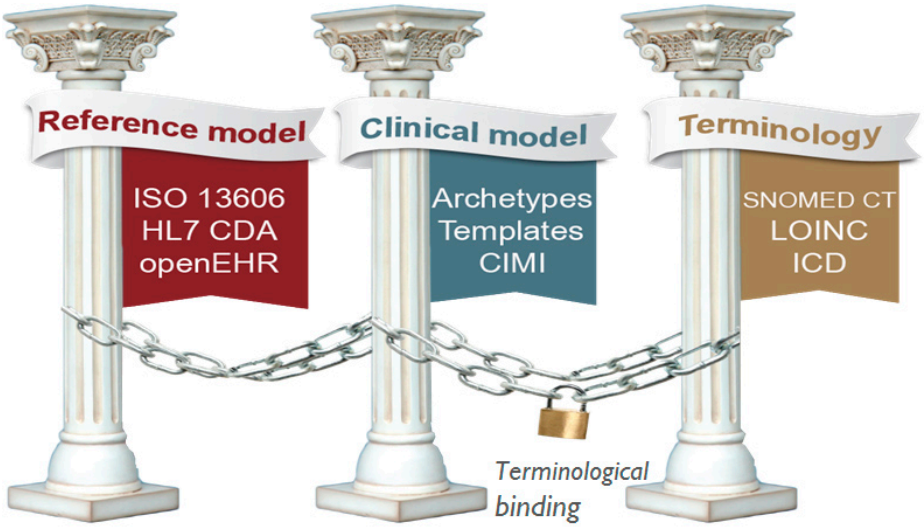

Figura 2. Los tres pilares sobre los que se asienta la interoperabilidad semántica

5. Representación de la información clínica

5.1 Historia Clínica Electrónica

Las organizaciones sanitarias, del mismo modo que cualquier empresa pública o privada, tiene la necesidad imperante de hacer uso de sistemas de información para la gestión de la información y los datos asociados a su quehacer diario. En ese sentido, la historia clínica (HC) es una de las fuentes de información más importantes en las instituciones médicas.

A grandes rasgos, la HC es el *conjunto de documentos que contienen los datos, valoraciones e informaciones sobre la situación y evolución clínica de un paciente a lo largo del proceso asistencial*. Asimismo, la HCE es *el conjunto de documentos y registros informáticos que contiene forma clara y concisa los datos, valoraciones e informaciones generados en cada uno de los procesos asistenciales a que se somete un paciente y en los que se recoge el estado de salud, la atención recibida y la evolución clínica de la persona*.

La HCE debe contener la información demográfica suficiente para identificar al paciente y la información clínica para apoyar el diagnóstico, justificar el tratamiento, documentar la evolución y los resultados de los tratamientos y promover la continuidad de la atención sanitaria. La información que recoge la HCE debe comprender los datos estrictamente necesarios y pertinentes, a fin de que de acuerdo con el principio de calidad o proporcionalidad al que se refiere el artículo 4 de la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, en ningún caso se recojan datos de las personas como pacientes que no resulten relevantes para facilitar su asistencia sanitaria ni aporten información veraz y actualizada acerca de su estado de salud.

La HCE tiene los siguientes usos:

- Asistencial: la misión principal de la historia clínica es proteger toda la información patográfica con objeto de prestar la mejor atención posible.
- Docente: con fines educativos.
- Investigación: tanto clínica como epidemiológica.
- Gestión clínica y planificación de recursos asistenciales.
- Jurídico legal: es testimonio documental de la asistencia prestada.
- Control de calidad asistencial.

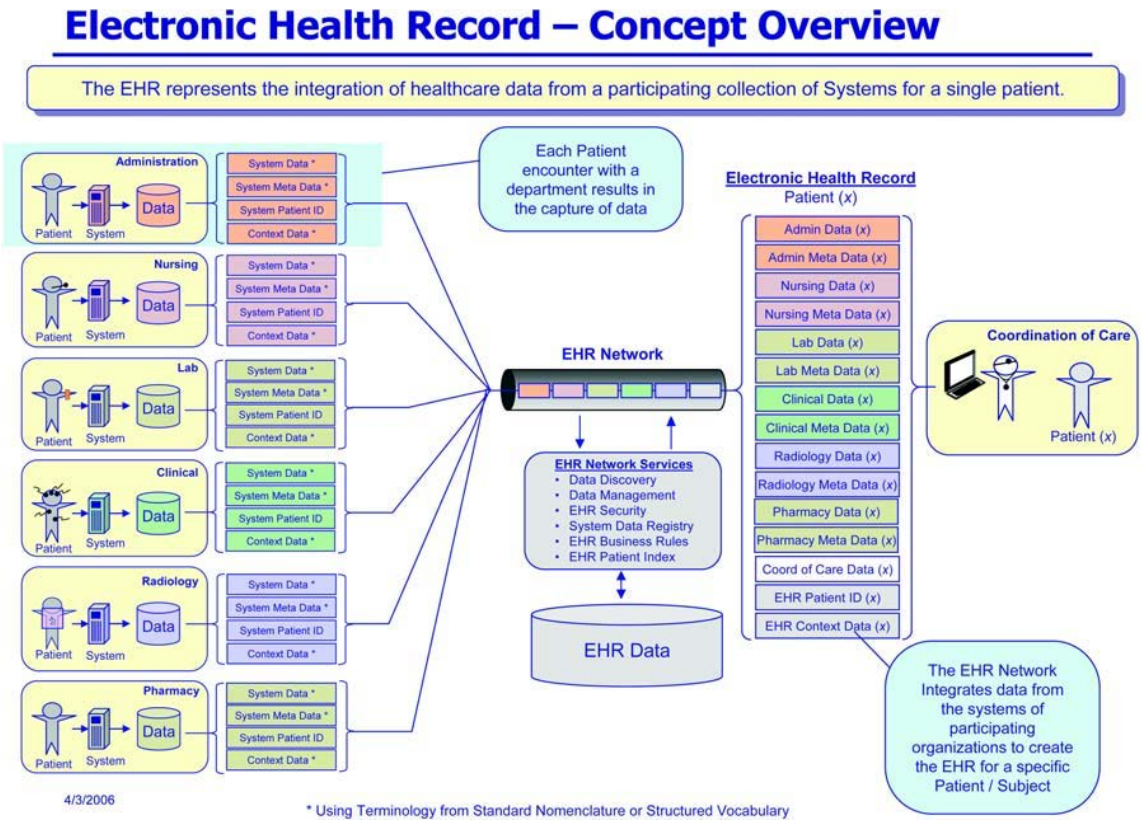

Figura 3. Vista general de la HCE y sus componentes (fuente: Johns Hopkins University)

A pesar del amplio consenso que existe sobre los beneficios de las HCE, su tasa de adopción es dispar en el mundo entero. Se observan muy buenas tasas de adopción en Australia, Holanda, el Reino Unido y Nueva Zelanda, así como en España y países nórdicos. En los Estados Unidos, la tasa de adopción era baja, pero ha mostrado un importante avance debido a la implementación de incentivos fiscales por parte del gobierno. Dichos incentivos son otorgados según criterios de uso significativo (meaningful use) de las funcionalidades de las HCE.

La historia clínica es una fuente de datos fundamental y una herramienta básica para la investigación biomédica, la formación de estudiantes y la educación médica continuada. Su implantación esta´ teniendo impactos en la investigación clínica, en la investigación farmacéutica (ensayos clínicos, farmacoepidemiologi´a) y en la investigación en salud pública (informe electrónico de casos, bases de datos poblacionales). En gran parte esto se debe a la creación de repositorios de datos de investigación que comienzan a estar integrados con la HCE de cada paciente, facilitando el desarrollo y la adopción de herramientas de soporte a la toma de decisiones o guías de práctica clínica que contribuyen al ejercicio de una medicina basada en evidencias.

5.2 Estándares de arquitecturas de HCE

La interoperabilidad semántica, según el Comité Europeo de Normalización (CEN), es el estado que existe entre dos aplicaciones cuando, en relación a una tarea específica, una de las aplicaciones es capaz de recibir datos de la otra y realizar satisfactoriamente dicha tarea, y todo ello sin que intervenga un operador externo. Dicho de otra manera, para que dos sistemas sean semánticamente interoperables, la información ha de fluir entre ellos de manera que el significado original se mantenga inalterado. Cada uno de estos sistemas ha de entender lo que recibe del otro y reaccionar acorde automáticamente. En el contexto sanitario, la interoperabilidad semántica es un requisito necesario para que la información dispersa de un paciente, generada en distintos lugares, distintos sistemas y en distintos instantes temporales, pueda ser compartida y esté a disposición de los profesionales clínicos siempre que sea necesaria e incluso pueda ser utilizada a efectos estadísticos o de investigación.

Para alcanzar un alto grado en interoperabilidad semántica entre los sistemas de información sanitarios es necesario el uso de estándares de HCE. El empleo de estándares de información es el único mecanismo que nos puede asegurar que sistemas diferentes sean capaces de interoperar semánticamente entre ellos. Cabe destacar que la adopción de estándares de información sanitaria no solo será de utilidad para la interoperabilidad semántica.

Los estándares son necesarios para ofrecer un lenguaje común de comunicación entre los distintos participantes, en cualquier acto socio-económico, aumentando la seguridad, disminuyendo los costos y favoreciendo el desarrollo de los mercados. En el caso de la tecnología, los estándares facilitan alcanzar la interoperabilidad entre sistemas, esto es, tener la capacidad de comunicarse e intercambiar procesos o datos.

Los estándares disponibles hoy en día están organizados en las siguientes cinco categorías:

1. *Estándares de mensajería e intercambio de datos:*

   Permiten el intercambio entre los sistemas y organizaciones en forma consistente, debido a que contienen instrucciones (o especificaciones) para el formato, los elementos de los datos y la estructura.

   Estándares comunes que incluyen la mensajería HL7 v2.x; El estándar EN13606 parte 5 (especifica las interfaces que deben cumplir los sistemas para poder comunicar extractos de HCE estandarizados); el estándar DICOM para las imágenes radiológicas; y el NCPD para las prescripciones electrónicas.

Organización	Siglas	Descripción	URL
Health Level Seven	HL7 V2.x HL7 V3	Mensajes para intercambiar datos demográficos, clínicos y administrativos.	hl7.org
Digital Imaging Communications in Medicine Committee	DICOM	Formatos para comunicar imágenes diagnósticas y datos asociados.	nema.org
Accredited Standards Committee X12	ASC X12	Mensajes para intercambiar tramitaciones, elegibilidad de pacientes y pagos de prestaciones.	X12.org
Institute of Electrical and Electronic Engineers Standard 1073	IEEE 1073	Mensajes para intercambiar datos con equipos de instrumentación biomédica.	standards.ieee.org

Tabla 1. Resumen de los estándares de intercambio de datos y mensajería

2. *Estándares de terminología:*

Básicamente son vocabularios que proveen códigos específicos para conceptos clínicos tales como enfermedades, listas de problemas, alergias, medicaciones y diagnósticos.

Ejemplos de estándares de terminología son LOINC para resultados de laboratorio; SNOMED CT para términos clínicos; y CIE para los diagnósticos médicos.

Organización	Siglas	Descripción	URL
International Classification of Diseases	CIE 9 CIE 10	Catálogo de diagnósticos y procedimientos para fines estadísticos, facturación, costes, tramitaciones, etc.	who.int/es/
Logical Observation Identifiers Names and Codes	LOINC	Orientado a órdenes médicas, peticiones de laboratorio y resultados.	loinc.org
International Health Terminology Standards Development Organisation Systematized Nomenclature of Medicine	SNOMED CT	Catálogo de conceptos biomédicos con descripciones , relaciones y gramática para construir expresiones clínicas.	ihtsdo.org/snomed-ct/
Unified Medical Language System	UMLS	Base de datos con más de un centenar de catálogos en múltiples idiomas y herramientas para mapear conceptos.	nlm.nih.gov/research/umls/

Tabla 2. Resumen de algunos estándares de terminología

3. *Estándares de documentos:*

Estos estándares definen qué tipo de información está incluida en un documento clínico, dónde puede ser hallada y cómo debe ser custodiada para garantizar su persistencia.

El CCR (Continuity of Care Record o Registro de cuidado continuo) es un estándar para la comunicación entre profesionales de la salud que incluye información de identificación de los actores que intervienen en la transferencia de pacientes, historia clínica, medicación concomitante, alergias y recomendaciones para el plan del cuidado en salud.

La norma EN13606, parte 1, permite representar cualquier documento e información clínica de la HCE así como la HCE al completo.

El estándar openEHR es un estándar abierto y describe la gestión y el almacenamiento de la información clínica mediante informes de HCE. openEHR es una fundación sin ánimo de lucro que se centra en la especificación de estándares en el ámbito de la informática médica.

Finalmente, HL7 CDA (Clinical Document Architecture) es un estándar diseñado para la representación, persistencia y comunicación de documentos clínicos.

4. *Estándares de aplicaciones:*

Se utilizan en la implementación de las reglas de negocio y su interacción con los sistemas de software. Facilitan la federación de identidades para autentificar usuarios y para consolidar vistas de información a través de múltiples bases de datos no integradas.

Los ejemplos incluyen:

Log-in único, que permite al usuario acceder a múltiples aplicaciones dentro del mismo ambiente; estándares que proveen una manera de ver la información en forma integradora a través de bases de datos. Un ejemplo es la gestión de identidades a través de sistemas de directorio tipo LDAP.

5. *Estándares de arquitectura:*

El estándar EN12967 (Health Informatics Service Architecture, HISA) es el marco para la construcción de todo el sistema de información sanitario.

Los objetivos principales de HISA son la definición de un modelo de información común para la construcción de servicios de un sistema de información sanitario y ofrecer una capa middleware independiente de los fabricantes que permita la interconexión de sistemas heterogéneos distribuidos.

5.3 Arquitecturas de HCE: el modelo dual

El propósito de las arquitecturas de HCE es transferir la HCE de los pacientes o bien un subconjunto de la misma. Para poder llevar esto a término es necesario utilizar uno o más sistemas físicos, ofreciendo la posibilidad de que su acceso sea multinacional o multinstitucional, sin olvidar que estos sistemas necesitan interactuar con otros servicios que proporcionan la terminología, conocimiento médico, seguridad y datos demográficos, entre otros.

En este sentido, es la representación de esta norma lo que la hace diferente de otras normas que no se adaptan al paso del tiempo, ya que los conjuntos de datos clínicos, las plantillas o la representación que se requiere por los distintos dominio sanitarios son diversos, complejos y cambian continuamente con el avance del trabajo y el conocimiento clínico. Además, permite representar cualquier tipo de estructura de datos de HCE de una manera consistente [1].

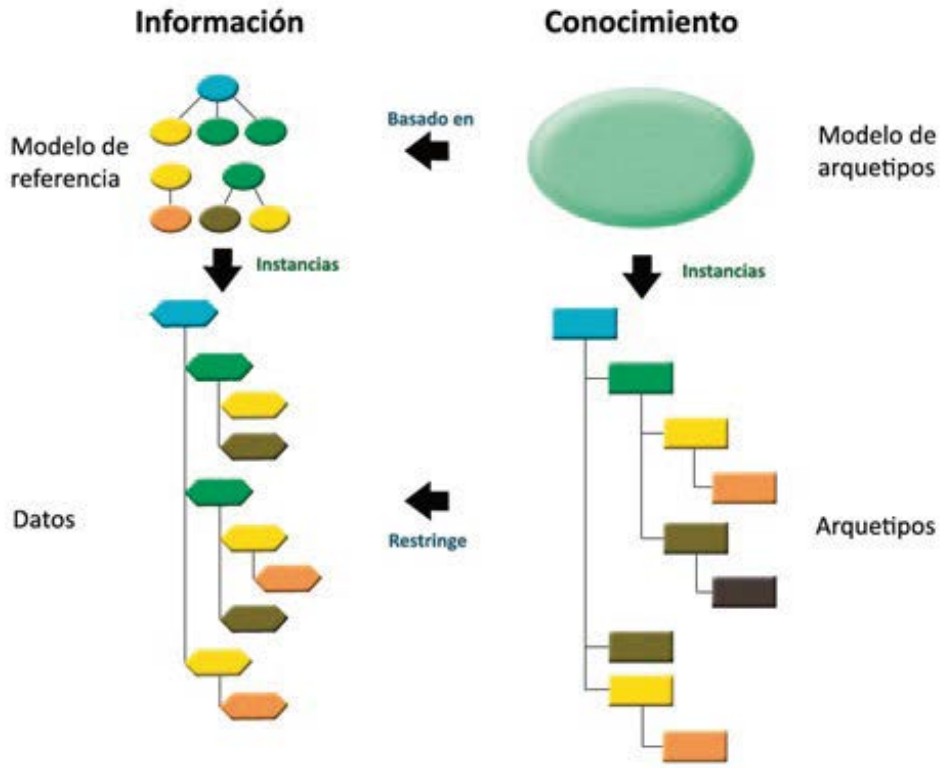

Figura 4. Diagrama del modelo dual

El modelo que solventa este problema es el llamado modelo dual. El modelo dual separa la parte que no varía con el tiempo, es decir, la información, de la parte que sí evoluciona, es decir, el conocimiento: La información no se modifica con el paso del tiempo. Son los datos clínicos de un paciente que se almacenan en el sistema de información, como pueda ser la medida del colesterol de un paciente o los datos relativos a un análisis de sangre en un momento concreto.

Ejemplo:

- **Información:** Juan García tiene una presión arterial de 150/100 mmHg a fecha 15 de enero de 2007.
- **Conocimiento:** La medida de la presión arterial tiene dos valores: sistólica y diastólica. Se miden en mmHg. Los rangos normales de presión arterial son entre 90/60 mmHg (en recién nacidos) hasta 135/85 mmHg, con un valor promedio de 120/80 mmHg.

Para la composición de estos datos se hace uso de los elementos y la organización del *modelo de referencia*. El conocimiento representa el conjunto de sucesos que forman un concepto, de forma que permite interpretar los datos. Este conocimiento varía y evoluciona con el paso del tiempo, y es lo que dota a este modelo de la capacidad de interoperar semánticamente y continuar siendo útil con el progreso del conocimiento, representado por el *modelo de arquetipos*.

El modelo de referencia especifica propiedades estables y genéricas de la HCE. Está formado por bloques de construcción genéricos de la HCE y define las maneras de combinar estos bloques para crear estructuras de datos más complejas. Además, define información contextual que debe acompañar a los datos para cumplir requisitos legales éticos y de proveniencia. Se implementa ad-hoc en esquemas de bases de datos o software por profesionales técnicos.

El modelo de arquetipos se ocupa de representar conceptos clínicos de una manera formal. Se enlazan con terminología clínica (i.e. enlace semántico y enlace de contenido o de valor). Son una combinación jerarquizada de estructuras del modelo de referencia restringida para ajustarse a la definición del concepto modelado. Los arquetipos son modelados por especialistas clínicos. El modelo de arquetipos es un marco que nos permite conjugar las entidades del modelo de referencia para crear estructuras con una mayor riqueza semántica. Un arquetipo es una definición formal que documenta la estructura de información asociada a un concepto clínico. Por ejemplo, el concepto de “Informe de alta”, “Medida de glucosa” o “Historia familiar”. Intuitivamente se podría interpretar un arquetipo como la plantilla de un documento clínico o una parte del mismo.

Algunas de las ventajas de esta separación de modelos son la independencia del software con los conceptos clínicos, ya que está condicionado por el modelo de referencia; los sistemas no necesitan ser modificados cuando los conceptos clínicos son creados o alterados y los clínicos pueden modelar conceptos sin depender de los técnicos.

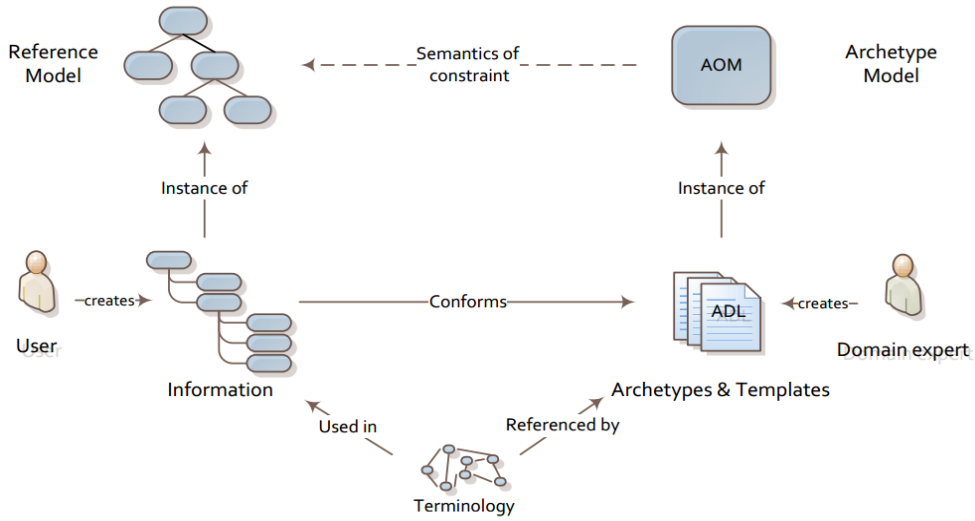

Figura 5. Esquema general del modelo dual

Los arquetipos se usan por los sistemas de información basados en un modelo dual para:

- la creación y validación de extractos de datos.
- la construcción de interfaces de usuario.
- la consulta flexible (semántica) de datos.
- la compartición de definiciones semánticas.

Cabe destacar que los arquetipos están pensados para que sean los propios profesionales sanitarios (que son los expertos en su dominio de conocimiento) los que los desarrollen utilizando para ello las herramientas adecuadas.

Desde el punto de vista de las estructuras de información, un arquetipo proporciona:

- Una estructura por defecto para la información de la HCE, basada en elementos del modelo de referencia.
- Restricciones en los valores, tipos, cardinalidad, etc. de dichos elementos que dan lugar a instancias válidas de datos.

Desde el punto de vista semántico o del conocimiento, un arquetipo proporciona:

- Una descripción semántica de alto nivel de la información asociada a conceptos clínicos, que puede ser procesada automáticamente por los sistemas de información sanitarios.
- Una independencia de los sistemas informáticos frente a los cambios en el dominio de conocimiento.

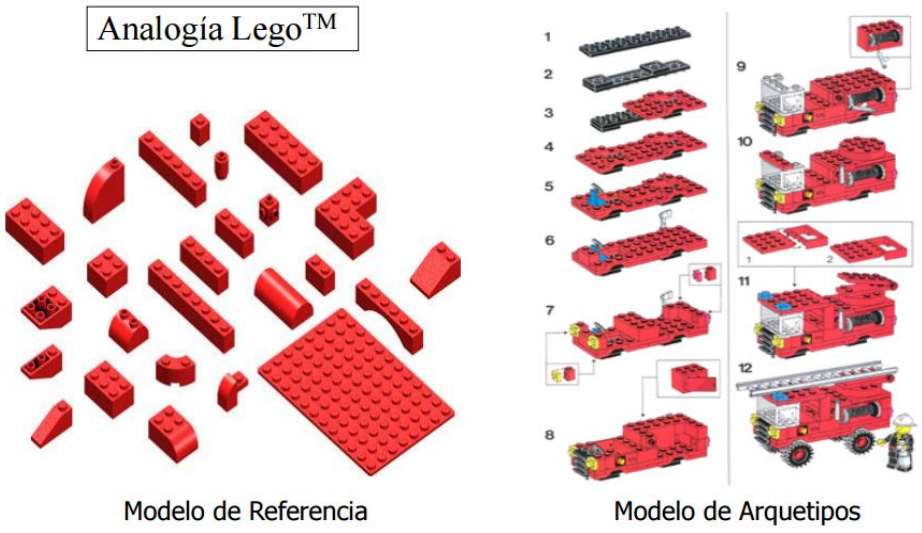

Figura 6. Analogía Lego para modelo de referencia y arquetipo

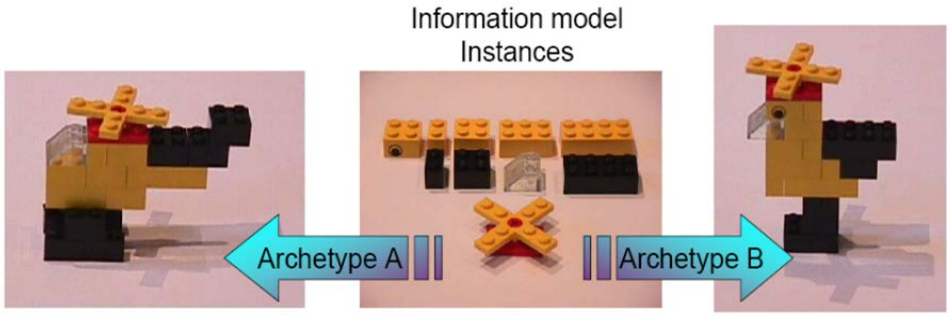

Figura 7. Analogía Lego para modelo de referencia y arquetipos

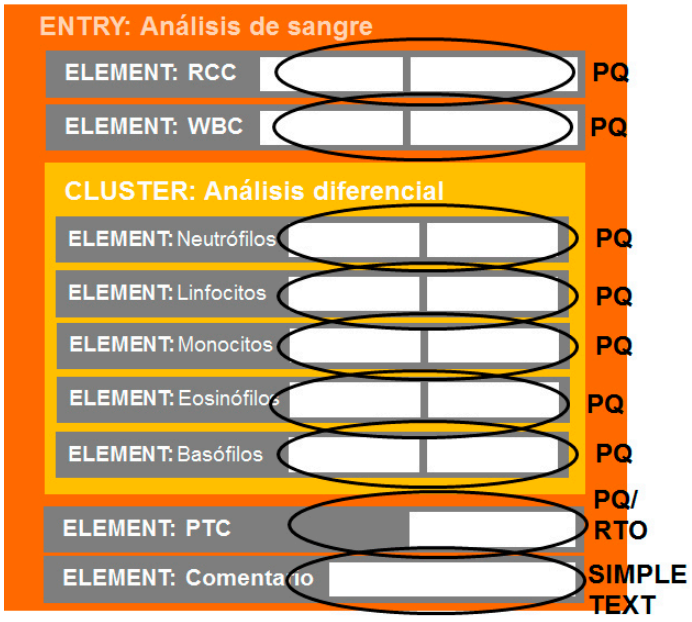

Figura 8. Ejemplo de arquetipo ISO 13606: Análisis de sangre

Los arquetipos se definen mediante el Archetype Definition Language (ADL) [2, 3]. La sintaxis del lenguaje ADL es independiente del modelo de referencia, de modo que gracias a este lenguaje es posible construir arquetipos ISO 13606, openEHR o HL7 CDA.

Estas son las secciones de un arquetipo (en lenguaje ADL):

- **Cabecera** con identificador de arquetipo
- **Identificador** del arquetipo **padre**, en caso de ser una especialización (opcional)
- Código del **concepto** que representa el arquetipo
- El **lenguaje** original del arquetipo
- Sección "**description**" con los metadatos del arquetipo (opcional)
- Sección "**definition**" con las restricciones formales del arquetipo
- Sección "**assertions**" conteniendo las reglas invariantes (opcional)
- Sección "**ontology**" con las definiciones de los términos en diferentes idiomas, terminologías y bindings.
- Sección "**revisión_history**" con el historial de cambios y auditoría (opcional).

6. Terminología SNOMED CT

Responde a las siglas Nomenclatura Sistematizada de Términos Clínicos. SNOMED International es la organización que se encarga de mantener y distribuir SNOMED CT. Lanzan una nueva versión cada 6 meses. Está considerada como la terminología más completa. Es escalable, es decir, ofrece la posibilidad de añadir conceptos nuevos a partir de unos mecanismos que se explican más adelante (i.e. expresiones pre- y post-coordinadas). Está traducida a varios idiomas: Inglés, Español, Danés, Holandés y Sueco. Contiene extensiones locales que contienen conceptos nuevos creados específicamente para cada extensión según los requerimientos. En España tenemos la extensión que desarrolla y mantiene el servicio de semántica del Servicio Nacional de Salud (SNS). El contenido de SNOMED CT está modelado/desarrollado por profesionales de la salud [4].

Figura 9. Distintas extensiones locales de SNOMED CT

SNOMED CT dispone de mapeos (i.e. correspondencias) con otros estándares terminológicos, como ICD-9, ICD-10 o LOINC. Se usa en más de 50 países.

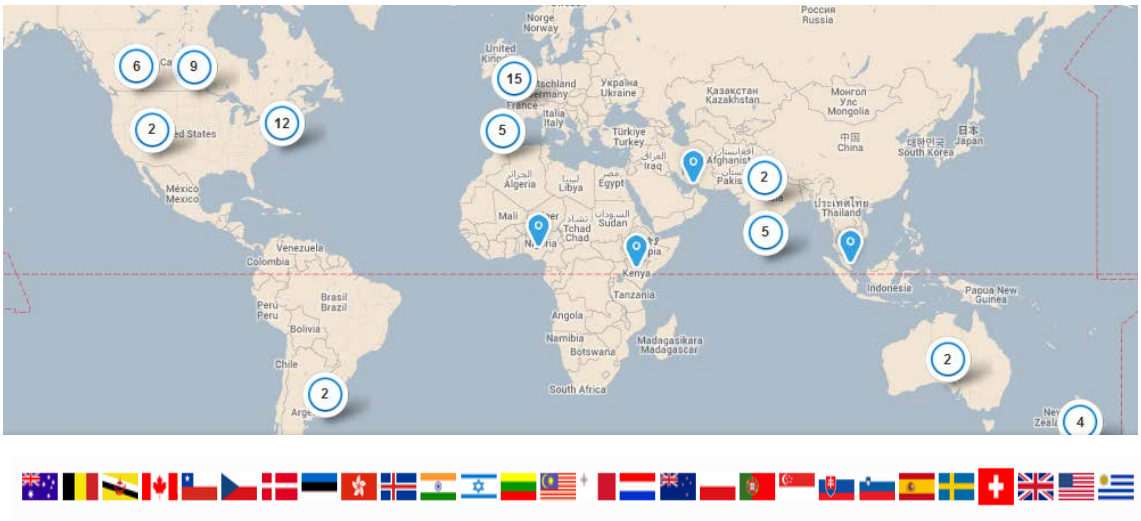

Figura 10. Mapa con los distintos países donde se usa de SNOMED CT

El *modelo lógico* de SNOMED CT define los componentes y la estructura que representan su contenido, concretamente: conceptos, descripciones y relaciones.

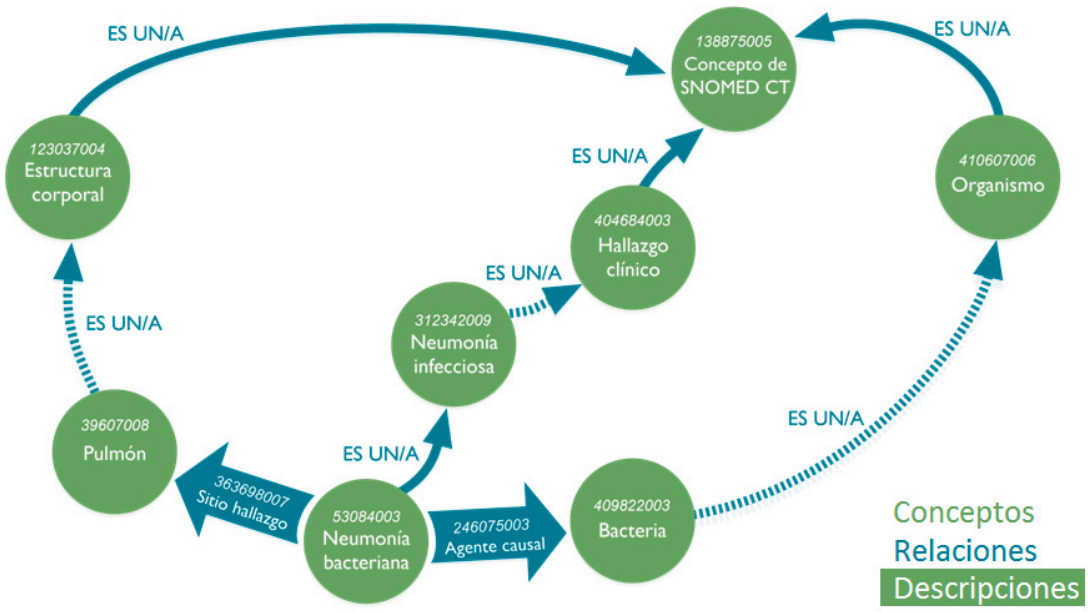

Figura 11. Fragmento de SNOMED CT con conceptos, relaciones y descripciones

Un *concepto* de SNOMED CT es un identificador numérico que representa un significado clínico único. Por ejemplo, 277170006, 217082002, 31978002, 43706004...

Una *descripción* es un texto que se asigna a un concepto para ser legible por el ser humano. Por ejemplo, 277170006 | edema de conducto auditivo |, 217082002 | caída accidental |, 31978002 | fractura de tibia |, 43706004 | vitamina C |... Cada concepto tiene asignada una descripción completa (i.e. Fully Specified Name - FSN) y varios sinónimos (uno preferido y el resto, aceptables). Asimismo, las propias descripciones también tienen un identificador asociado.

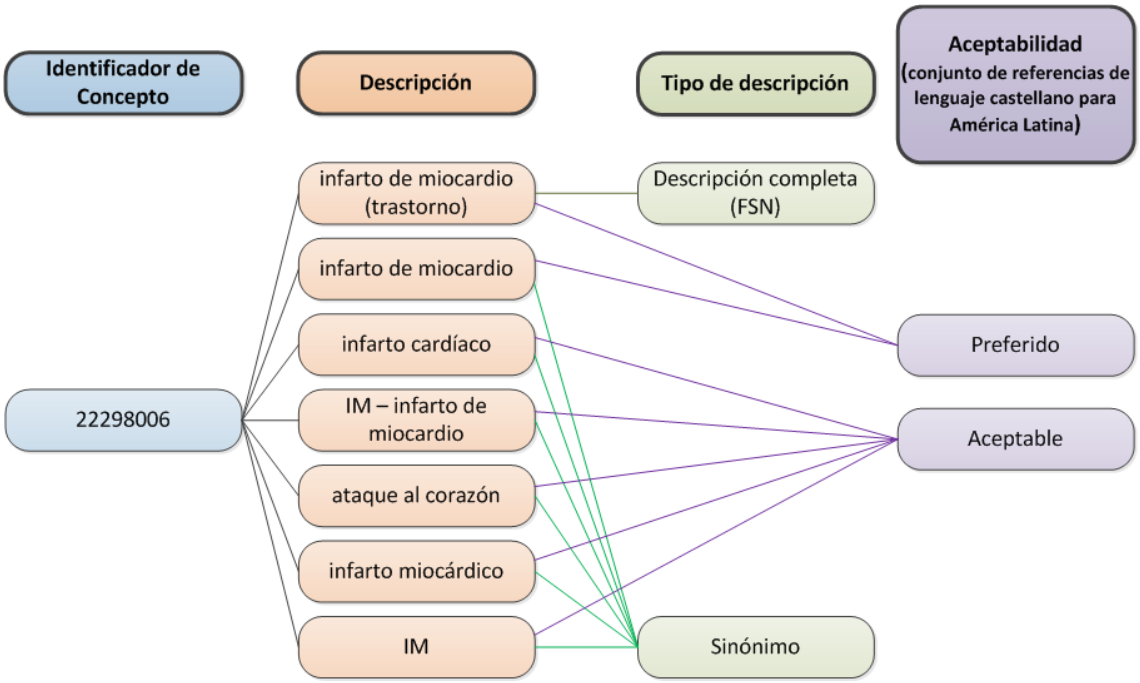

Figura 12. Ejemplo de concepto y sus descripciones asociadas, tanto la completa como los sinónimos (el preferido y los aceptables)

Una *relación* es una asociación entre dos conceptos. Hay dos tipos: de subtipo (i.e. es un/una) y de atributo. Esto es lo que hace que SNOMED CT sea tan potente, las relaciones. Es lo que hace que SNOMED CT sea una ontología. Los conceptos no son meras listas (como ocurre en las clasificaciones como CIE), sino que están conectados unos con otros. La relación es un/una define jerarquías, especializaciones. En SNOMED CT un concepto puede ser especialización de más de uno (i.e. polijerarquías). Hablamos de concepto padre y concepto hijo. Cuanto más profundo esté un concepto, mayor será su nivel de granularidad o detalle. Y cuanto más alto esté en la jerarquía, será un concepto más general o agrupador.

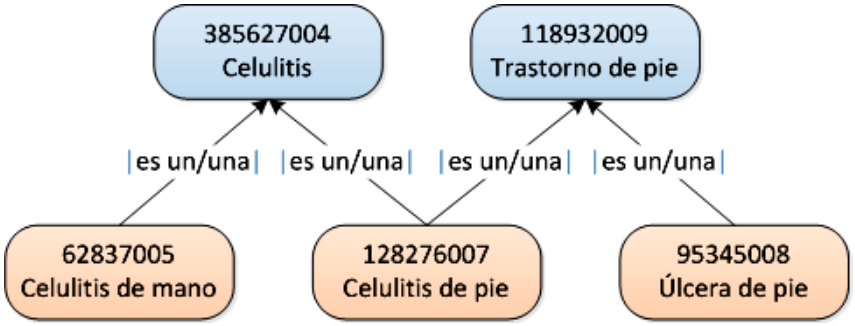

Figura 13. Ejemplo de dos conceptos y sus subtipos (nótese la polijerarquía)

Por su parte, las relaciones de atributo sirven para definir un concepto. Hay 40 tipos de atributos aproximadamente. Por ejemplo, lateralidad, severidad, tiene como principio activo a, prioridad...

En SNOMED CT los conceptos tienen asociados una *definición lógica*, formada por sus relaciones de subtipo y sus relaciones de atributo.

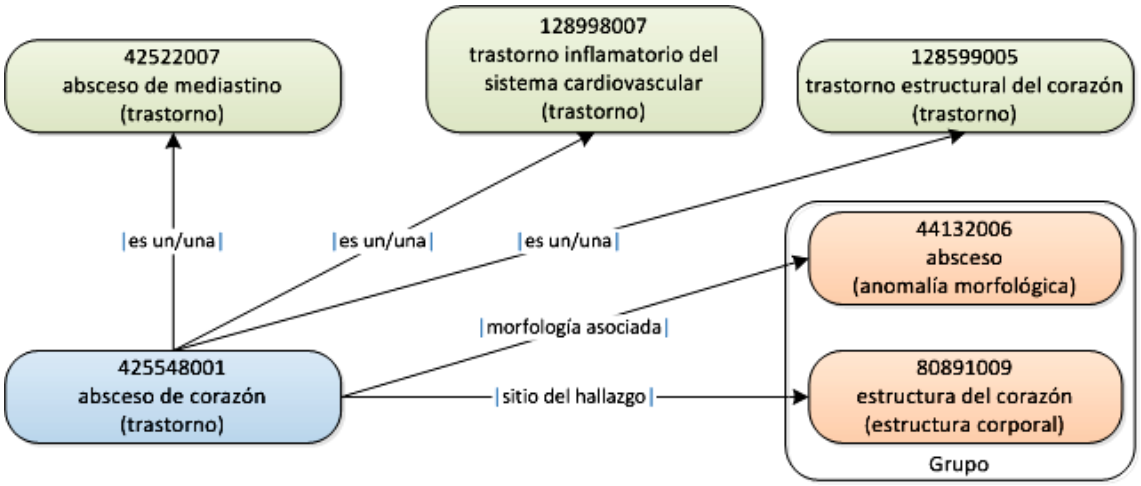

Figura 14. Definición lógica del concepto 425548001 |absceso de corazón|

Internamente, y vistas las relaciones que presenta SNOMED CT, es fácil vislumbrar que la estructura interna de SNOMED CT es la de un grafo. Concretamente se trata de un grafo dirigido y acíclico.

Figura 15. Distintos tipos de grafos

SNOMED CT se compone de una serie de jerarquías principales o de alto nivel (i.e. top level hierarchies). Aproximadamente existen 19 jerarquías principales, según la versión, entre ellas: procedimientos, hallazgos clínicos, organismos, sustancias, estructuras corporales, etc.

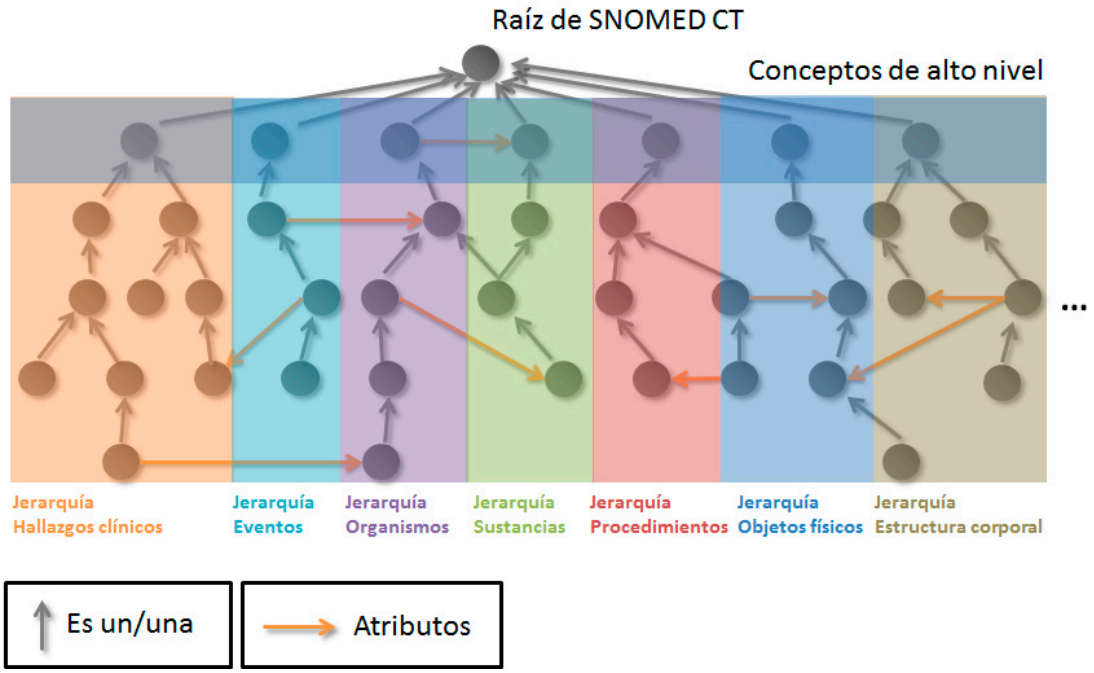

Figura 16. Algunas de las jerarquías principales de SNOMED CT

SNOMED CT por aproximadamente 300.000 conceptos activos (más de 400.000 entre activos e inactivos). Y más de un millón y medio de relaciones.

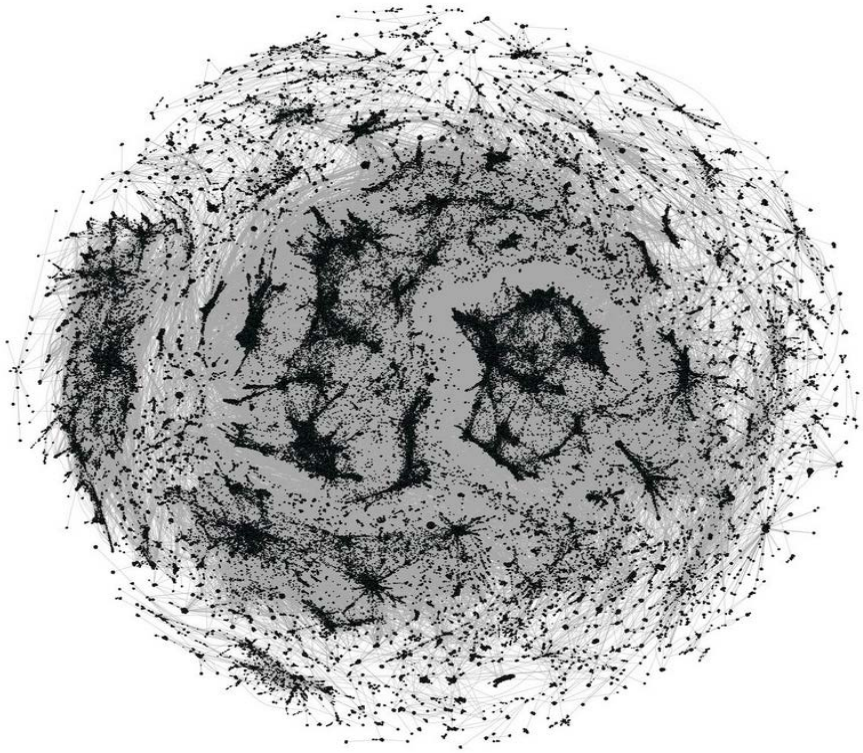

Figura 17. Este es el aspecto que presenta SNOMED CT, donde los puntos negros son los conceptos y las líneas grises son las relaciones de subtipo. Obsérvese que las jerarquías principales corresponden a las zonas más oscuras

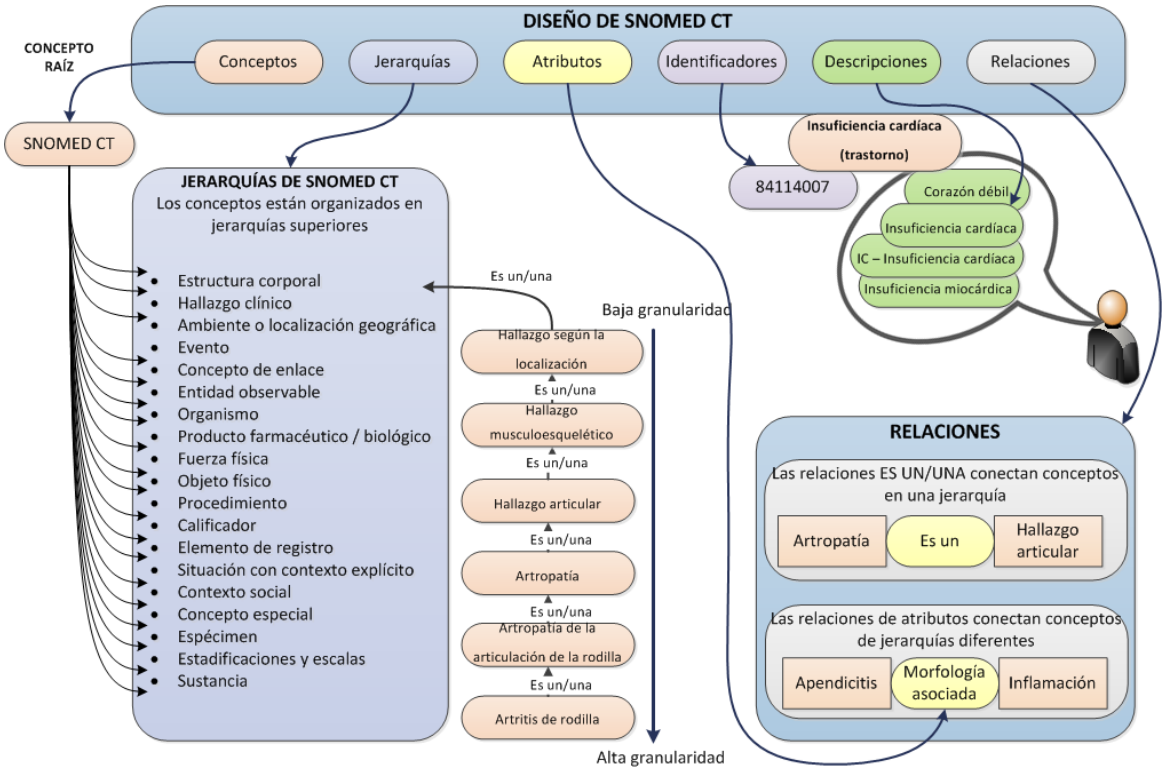

Figura 18. Vista general del modelo lógico de SNOMED CT

Atendiendo a su nivel de definición (i.e. a sus características definitorias), existen dos tipos de conceptos en SNOMED CT: los conceptos *primitivos* y los conceptos *totalmente definidos*.

Un concepto es totalmente definido si sus características definitorias son suficientes para distinguir su significado de otros conceptos similares. Por ejemplo, esta es la definición lógica del concepto 2704003 |enfermedad aguda|, que es totalmente definido:

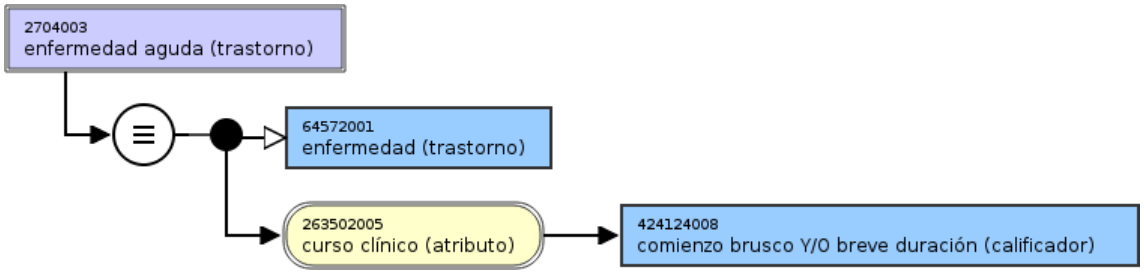

Figura 19. Definición lógica del concepto 2704003 |enfermedad aguda|

Cualquier concepto de SNOMED CT que incluya estas relaciones en su definición lógica será un descendiente (i.e. subtipo) de 2704003 |enfermedad aguda|.

En contraposición a los conceptos totalmente definidos están los conceptos primitivos, cuyas características definitorias no son suficientes para distinguir su significado de otros conceptos. Por ejemplo, los conceptos 64572001 |enfermedad| y 69449002 |acción de un fármaco| son conceptos primitivos.

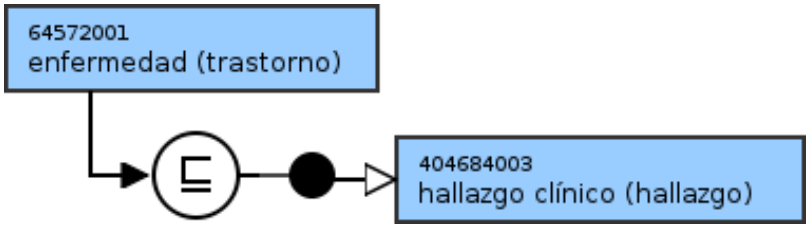

Figura 20. Definición lógica del concepto 64572001 |enfermedad|

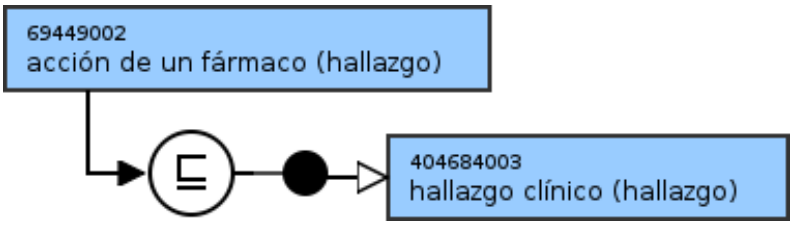

Figura 21. Definición lógica del concepto 69449002 |acción de un fármaco|

En SNOMED CT existen los denominados *conjuntos de referencias*. Los conjuntos de referencias: son una manera estándar de personalizar y mejorar el contenido de SNOMED CT para ser utilizado en un país, idioma, especialidad o contexto particular. Por ejemplo, subconjuntos de idiomas y dialectos, subconjuntos de mapeos y subconjuntos de conceptos.

El *modelo conceptual* de SNOMED CT son un conjunto de reglas que especifican el modo en que dos conceptos pueden relacionarse mediante relaciones de atributo. Dicho de otra manera, son las reglas que establecen el modo en que se relacionan los conceptos en la terminología. El modelo conceptual establece el dominio (i.e. concepto/s origen de la relación), el atributo y el rango (i.e. concepto/s destino de la relación). O dicho de otra manera, el dominio es la jerarquía/s donde un determinado atributo puede ser aplicado. Por su parte, el rango es la jerarquía/s permitidas como valores para un determinado atributo. El modelo conceptual es el mecanismo que marca la semántica de SNOMED CT.

Figura 22. Partes del modelo conceptual de SNOMED CT

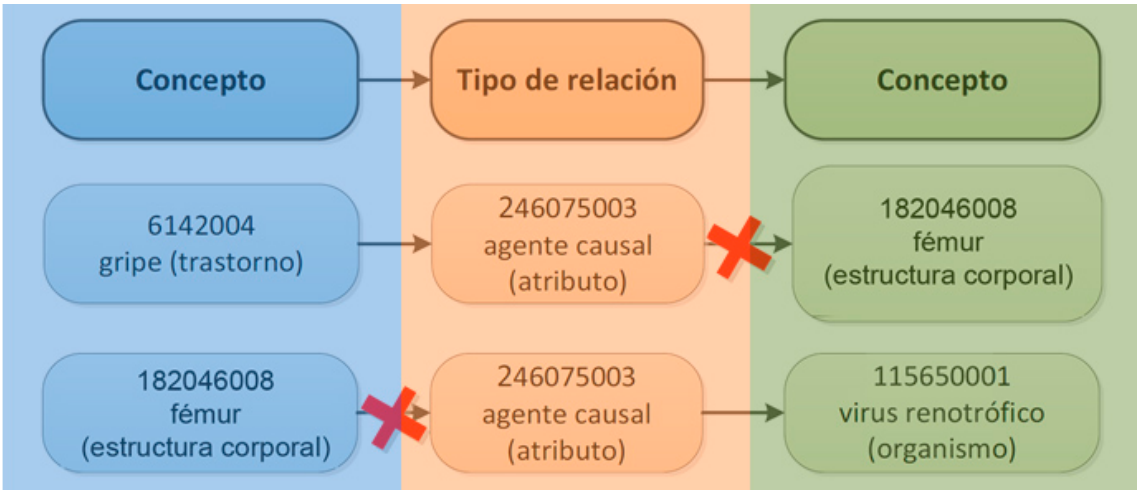

Figura 23. Ejemplos de relaciones correctas e incorrectas según el modelo conceptual

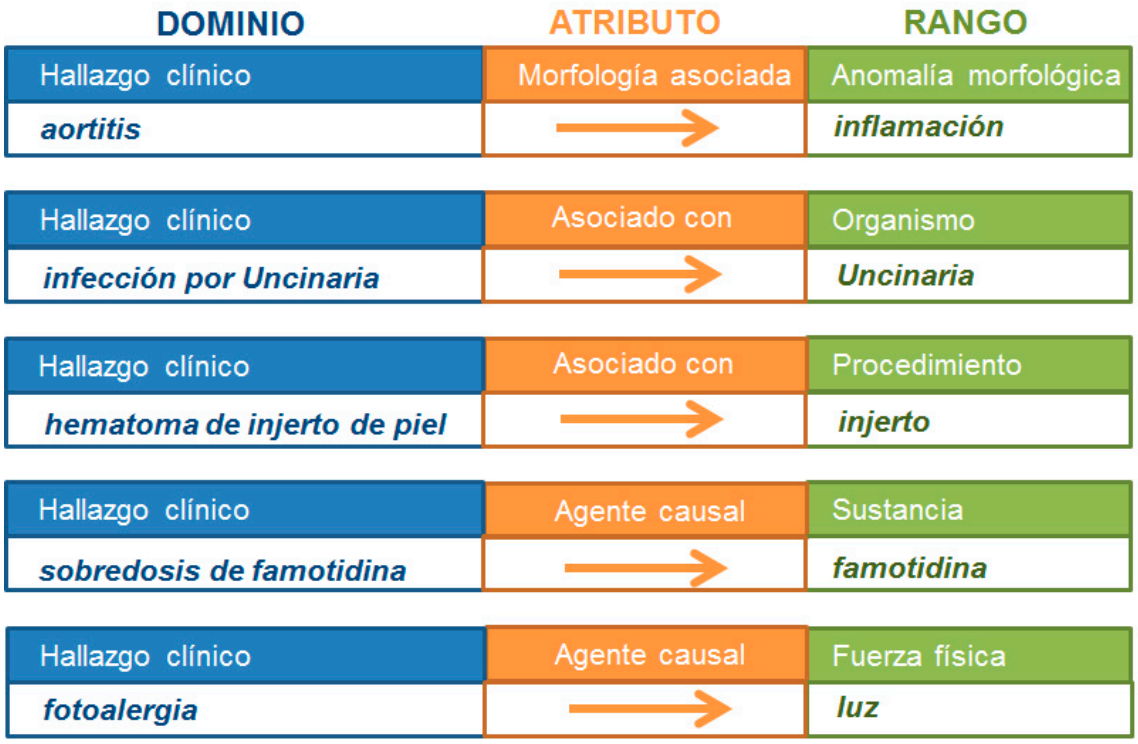

Figura 24. Ejemplos de relaciones correctas según el modelo conceptual

El *modelo conceptual procesable por computador* o MRCM (del inglés, Machine Readable Concept Model) representa las reglas del modelo conceptual de una manera que pueden ser leídas por un computador y aplicadas para validar que las definiciones de los conceptos y expresiones cumplen con estas reglas. El MRCM puede ser utilizado para una variedad de propósitos, incluyendo la creación y validación de conceptos de SNOMED CT, expresiones, restricciones de expresiones y consultas, procesamiento de

lenguaje natural (NLP) y enlaces terminológicos con modelos de información para dar soporte a la consulta y la interoperabilidad (todos estos conceptos se presentan más adelante en el presente trabajo). Actualmente el MRCM está siendo desarrollado por la SNOMED International y se halla en proceso de revisión de su primera versión.

En SNOMED CT, una *expresión* es un conjunto de conceptos que, sintácticamente combinados, representan un conocimiento clínico con el nivel de detalle requerido. Existen dos tipos: por una parte, las *expresiones pre-coordinadas*, formadas por un solo identificador (i.e. un concepto), y, por otra, las *expresiones post-coordinadas*, formadas por más de un identificador.

Por ejemplo, el concepto 174041007 |apendicectomía laparoscópica de emergencia| es una expresión pre-coordinada, cuya definición lógica es esta:

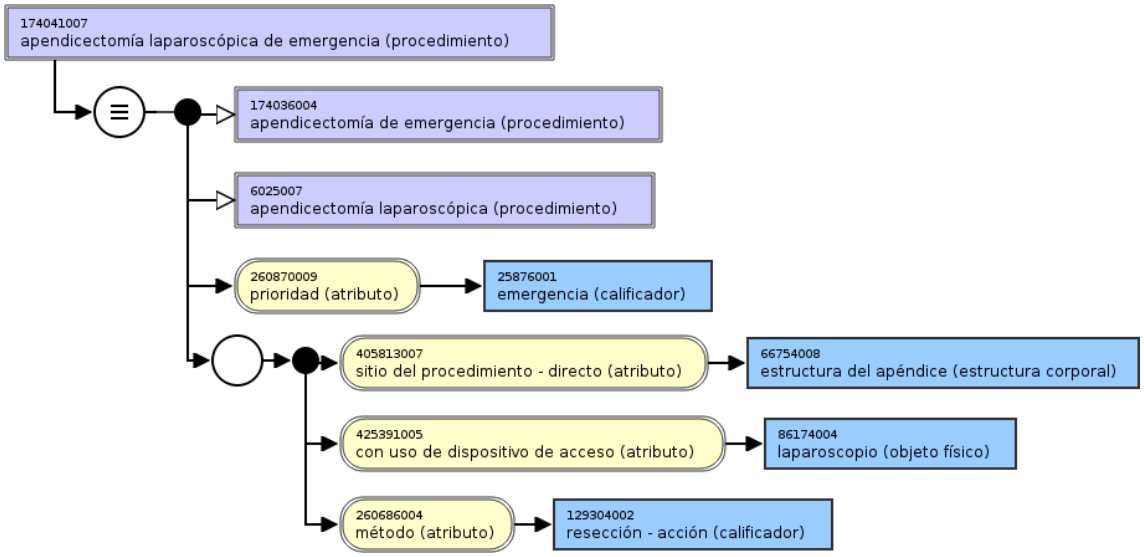

Figura 25. Definición lógica del concepto 174041007 |apendicectomía laparoscópica de emergencia|

El concepto pre-coordinado 174041007 |apendicectomía laparoscópica de emergencia| ya tiene el nivel de detalle deseado. Pero ¿qué pasaría si este concepto no existiera en SNOMED CT? Podríamos construirlo gracias al mecanismo de la post-coordinación. Para post-coordinar es necesario tener en cuenta las reglas del modelo conceptual de SNOMED CT para que la expresión tenga sentido (i.e. sea semánticamente válida). Esta sería la expresión post-coordinada equivalente:

80146002 |apendicectomía| :

260870009 |prioridad| = 25876001 |emergencia|,

425391005 |dispositivo de acceso| = 86174004 |laparoscopio|

Conviene subrayar que las expresiones post-coordinadas se construyen a partir de la *gramática composicional* [5]. La expresión post-coordinada anterior define una apendicectomía cuya prioridad es emergencia y el dispositivo de acceso es un laparoscopio.

El siguiente ejemplo también define una expresión post-coordinada. Concretamente se trata de una quemadura de piel con una morfología asociada de quemadura de tercer grado causada por agua caliente en el dedo índice de la mano izquierda:

284196006 |quemadura de piel| :

116676008 |morfología asociada| = 80247002 |tercer grado|,

246075003 |agente causal| = 47448006 |agua caliente|,

363698007 |sitio del hallazgo| = 83738005 |dedo índice|,

272741003 |lateralidad| = 7771000 |lado izquierdo|

A continuación se muestra el modelo lógico de la gramática composicional.

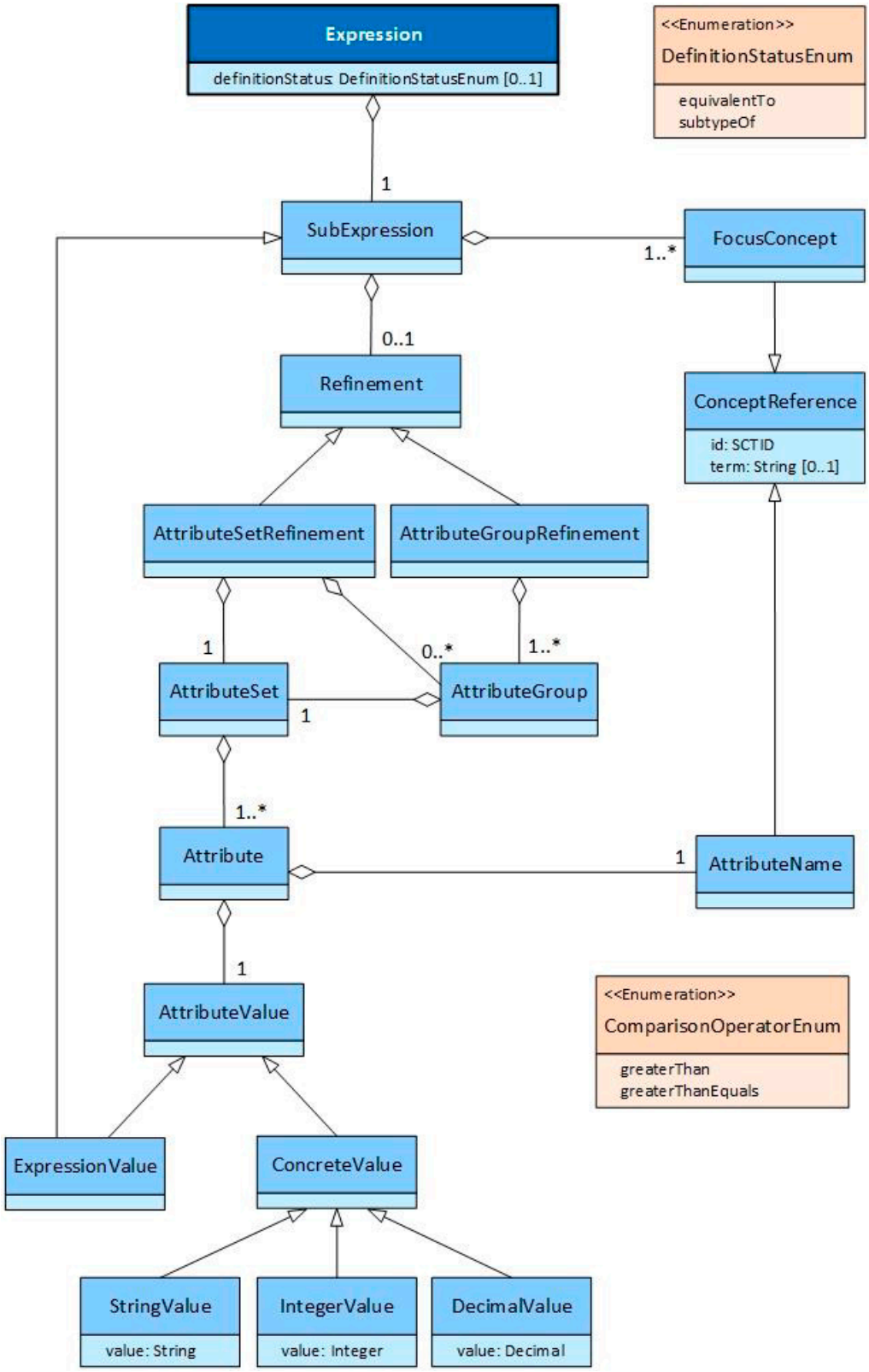

Figura 26. Modelo lógico de la gramática composicional de SNOMED CT

Y su sintaxis definida en ABNF.

```
expression = ws [definitionStatus ws] subExpression ws subExpression =
focusConcept [ws ":" ws refinement]
definitionStatus = equivalentTo / subtypeOf equivalentTo = "===" subtypeOf =
"<<<"
focusConcept = conceptReference *(ws "+" ws conceptReference) conceptReference
= conceptId [ws "|" ws term ws "|"]
conceptId = sctId
term = nonwsNonPipe *( *SP nonwsNonPipe )
refinement = (attributeSet / attributeGroup) *( ws [","  ws] attributeGroup )
attributeGroup = "{" ws attributeSet ws "}" attributeSet = attribute *(ws "," ws
attribute) attribute = attributeName ws "=" ws attributeValue
attributeName = conceptReference attributeValue = expressionValue / QM
stringValue QM / "#" numericValue expressionValue = conceptReference / "(" ws
subExpression ws ")"
stringValue = 1*(anyNonEscapedChar / escapedChar)
numericValue = decimalValue / integerValue
integerValue = (["-"/"+"] digitNonZero *digit ) / zero decimalValue = integerValue
"." 1*digit
sctId = digitNonZero 5*17( digit )
ws = *( SP / HTAB / CR / LF ) ; optional white space
SP = %x20 ; space
HTAB = %x09 ; tab
CR = %x0D ; carriage return
LF = %x0A ; line feed
QM = %x22 ; quotation mark
BS = %x5C ; back slash
digit = %x30-39
zero = %x30
digitNonZero = %x31-39
nonwsNonPipe = %x21-7B / %x7D-7E / UTF8-2 / UTF8-3 / UTF8-4
anyNonEscapedChar = HTAB / CR / LF / %x20-21 / %x23-5B / %x5D-7E / UTF8-2 /
UTF8-3 / UTF8-4
escapedChar = BS QM / BS
UTF8-2 = %xC2-DF UTF8-tail
```

UTF8-3 = %xE0 %xA0-BF UTF8-tail / %xE1-EC 2(UTF8-tail) / %xED %x80-9F UTF8-tail / %xEE-EF 2(UTF8-tail)

UTF8-4 = %xF0 %x90-BF 2(UTF8-tail) / %xF1-F3 3(UTF8-tail) / %xF4 %x80-8F 2(UTF8-tail)

UTF8-tail = %x80-BF

Las *restricciones sobre expresiones* de SNOMED CT consisten en añadir operadores de restricción y conjuntistas a las expresiones para definir subconjuntos de conceptos (i.e. de conocimientos clínicos) en lugar de uno solo (i.e. post-coordinación). Para definir estas restricciones se utiliza el recientemente creado por SNOMED International, Lenguaje de Restricciones de Expresiones de SNOMED CT [6]. Dada la importancia de este lenguaje para definir subconjuntos de conceptos, susceptibles de ser utilizados en el enlace de contenido o de valor entre modelos de información, tales como arquetipos, y SNOMED CT, en aras de obtener un nivel alto en interoperabilidad semántica, y dado que es el lenguaje al que se le da soporte en el motor de ejecución presentado en este trabajo final, le dedicamos el siguiente punto.

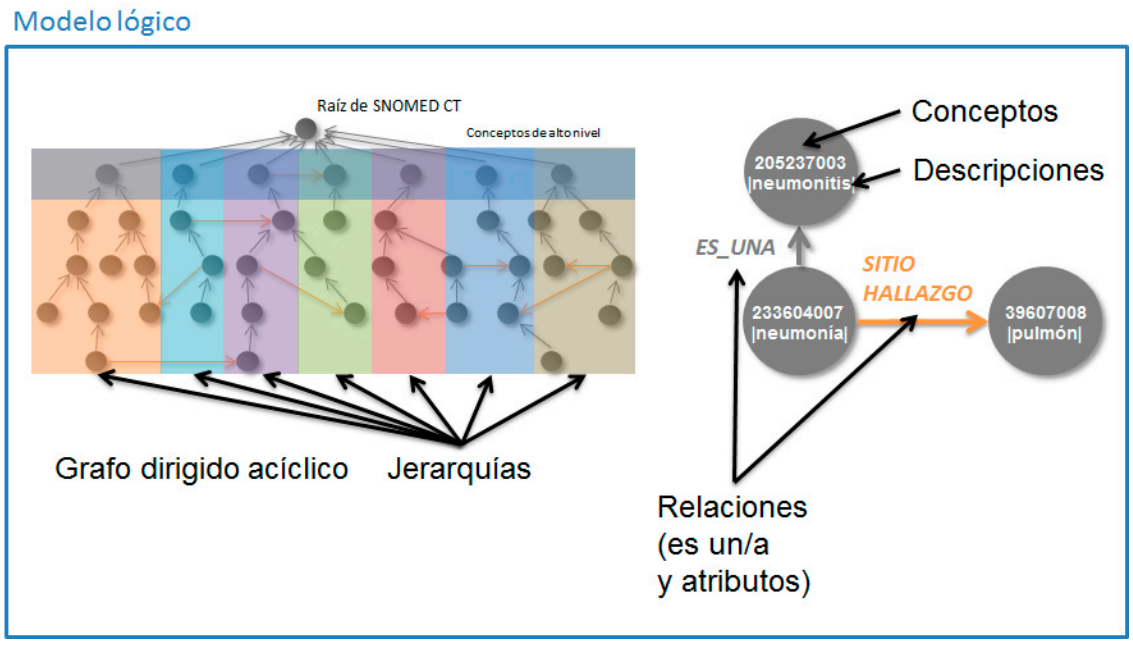

Figura 27. Resumen del modelo lógico de SNOMED CT

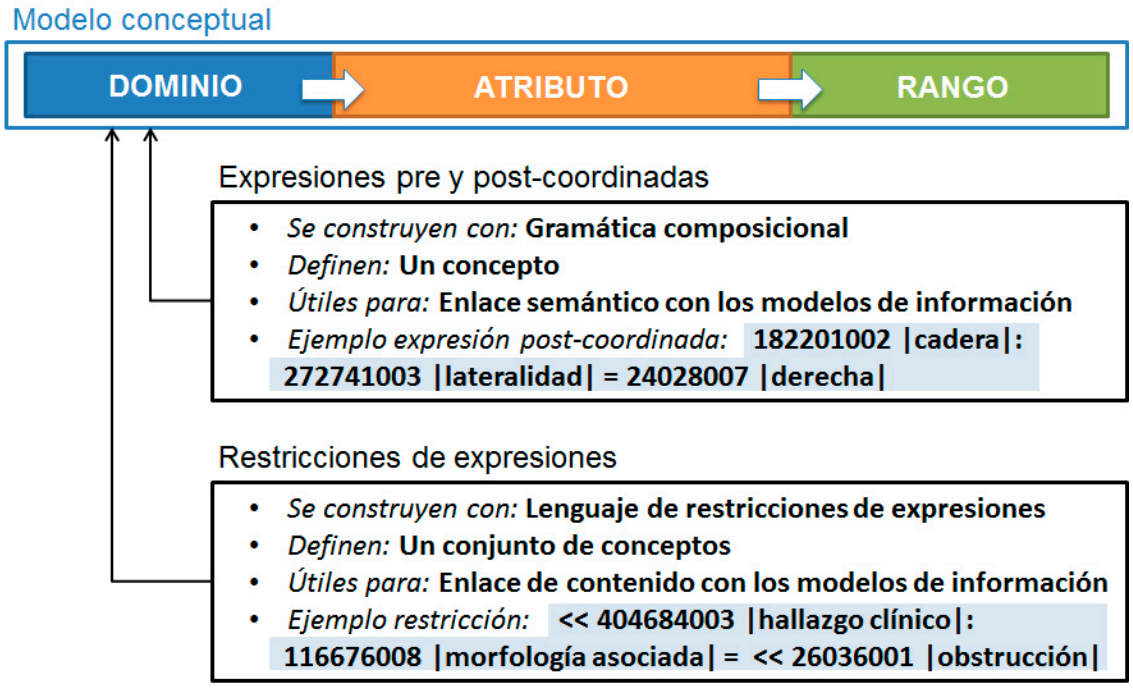

Figura 28. Resumen del modelo conceptual, expresiones y restricciones de expresiones de SNOMED CT

7. Enlace terminológico

Tal y como se ha visto en el punto anterior, para conseguir un alto grado de interoperabilidad semántica es imprescindible llevar a cabo un enlace terminológico entre los modelos de información y las terminologías.

Existen dos tipos de enlace terminológico:

- el *enlace semántico* [7,8,9,10], que proporciona significado inequívoco a las estructuras de información contenidas en el modelo de información mediante un enlace entre un elemento del modelo de información y un término pre o post-coordinado de la terminología (por ejemplo, el elemento *diabetes mellitus* de un arquetipo corresponde al concepto *73211009 | diabetes mellitus (trastorno)* | de SNOMED CT).

- el *enlace de contenido* (o enlace de valor) [1], que restringe el conjunto de posibles valores codificados de la terminología, susceptibles de ser asociados a un elemento del modelo de información (por ejemplo, el elemento *procedimiento quirúrgico* de un arquetipo está asociado al subconjunto de los procedimientos quirúrgicos de SNOMED CT susceptibles de ser seleccionados en dicho elemento, como *110468005 | cirugía ambulatoria (procedimiento) |*, *711364006 | cirugía con asistencia robótica (procedimiento) |*, *56306000 | cirugía estética (procedimiento) |*, y hasta un total de unos 20.000 procedimientos quirúrgicos).

Para lograr el enlace semántico basta con definir el enlace dentro del propio modelo de información clínico detallado entre el elemento en cuestión y el concepto terminológico. Si bien en la literatura científica se han abordado estrategias para crear enlaces semánticos de manera manual o semiautomática, el enlace de contenido no ha sido estudiado en profundidad.

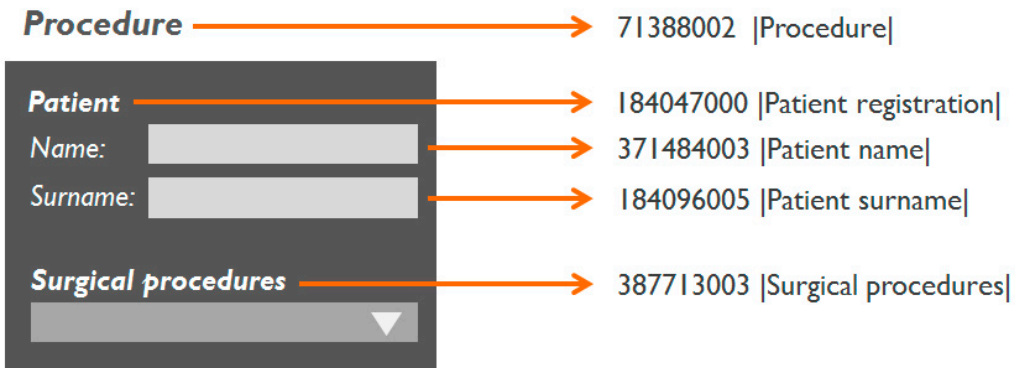

Figura 29. Ejemplo de enlace semántico entre un formulario con datos del paciente y los procedimientos quirúrgicos aplicados, y conceptos de SNOMED CT

Por su parte, para acometer el enlace de contenido, una primera opción es definir los subconjuntos de manera extensional (i.e. listando los elementos del subconjunto). Pero

esta aproximación tiene algunos problemas, como por ejemplo la posibilidad de que algún código cambie en el futuro, lo que conllevaría a tener códigos erróneos en los subconjuntos; o la disminución en la eficiencia y el elevado coste que tiene manejar subconjuntos grandes (recordemos que las terminologías médicas pueden llegar a contener cientos de miles de conceptos cínicos).

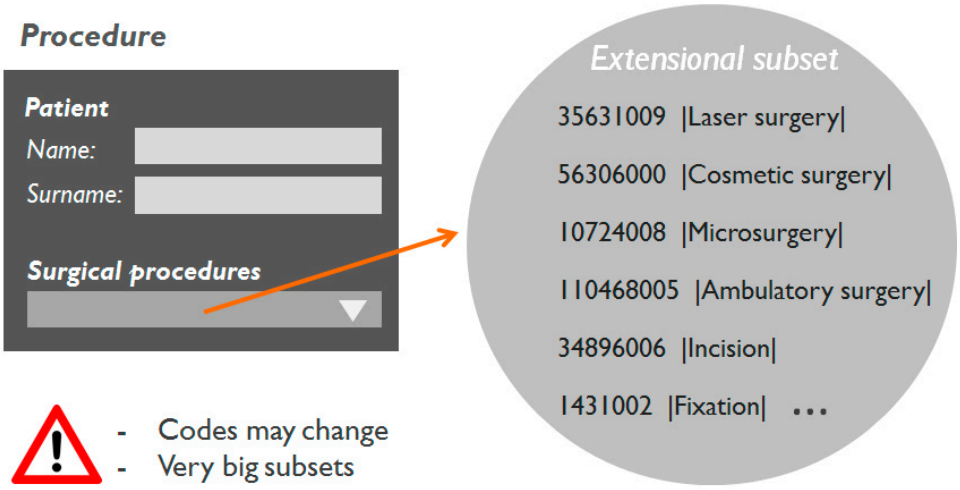

Figura 30. Ejemplo de enlace de contenido entre un formulario con datos del paciente y los procedimientos quirúrgicos aplicados, y un subconjunto extensional de SNOMED CT

Por ello, la mejor opción para llevar a cabo el enlace de contenido requiere un mecanismo de definición de subconjuntos de conceptos terminológicos, como el recientemente creado Lenguaje de Restricciones de Expresiones de SNOMED CT por parte de la organización de desarrollo de estándares terminológicos SNOMED International. Gracias a este lenguaje es posible definir los subconjuntos de conceptos médicos de manera intensional, es decir, mediante una expresión susceptible de ser evaluada para calcular los conceptos que forman el subconjunto (i.e. por comprensión).

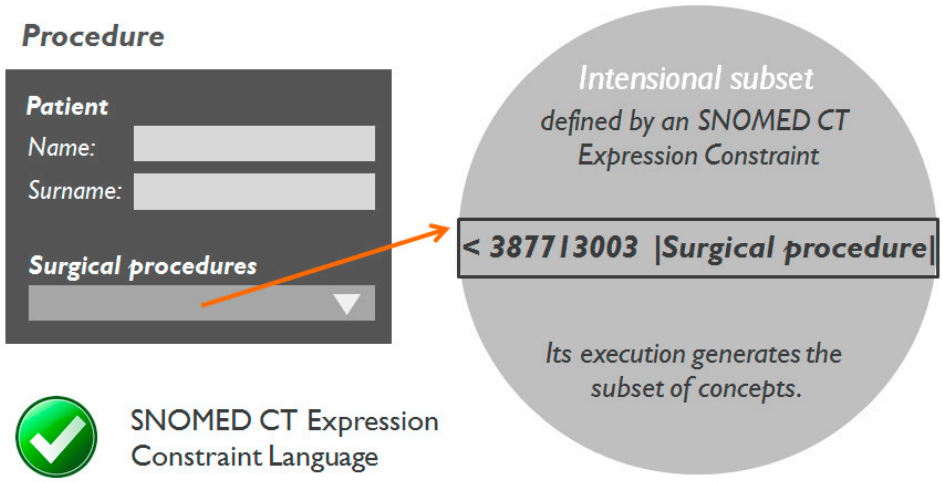

Figura 31. Ejemplo de enlace de contenido entre un formulario con datos del paciente y los procedimientos quirúrgicos aplicados, y un subconjunto intensional de SNOMED CT

8. Materiales y métodos

8.1 Lenguaje de Restricciones de Expresiones de SNOMED CT

El *Lenguaje de Restricciones de Expresiones SNOMED CT* es una sintaxis formal para representar restricciones de expresiones de SNOMED CT. Las restricciones de expresiones son reglas computables para definir subconjuntos delimitados de conocimientos clínicos representados por expresiones pre-coordinadas o post-coordinadas. Las restricciones de expresiones pueden ser utilizadas para restringir los valores válidos de un elemento de datos en una HCE, como la definición intensional (i.e. por comprensión) de un conjuntos de referencias de conceptos, como una consulta procesable por ordenador que identifica un conjunto de expresiones, o como una restricción que restringe el rango de un atributo definido en el modelo de conceptos de SNOMED CT.

Es importante recordar que la gramática composicional de SNOMED CT se emplea para definir conceptos post-coordinados (i.e. conceptos que no existen en la terminología pero que pueden ser requeridos en un momento dado), mientras que el lenguaje de restricciones se utiliza para definir subconjuntos de conceptos pre-coordinados, es decir, conceptos que existen como tales en la terminología.

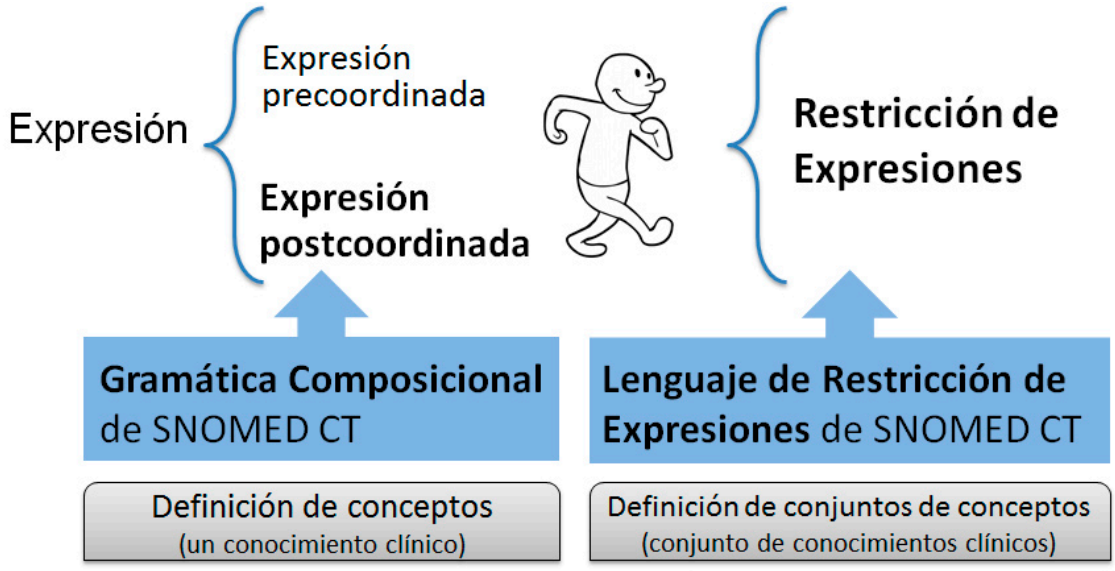

Figura 32. Comparación entre la gramática composicional y el lenguaje de restricciones

El lenguaje de restricciones de expresiones de SNOMED CT debe dar soporte a las capacidades que se recogen en la siguiente tabla.

Función	**Detalles**
Referencia a concepto	Capacidad para referenciar conceptos pre-coordinados de SNOMED CT usando su identificador y una descripción opcional en lenguaje natural.
Jerarquías de conceptos	Capacidad para seleccionar un conjunto de conceptos, como son los descendientes de un concepto, los descendientes y el propio concepto, los ascendientes y los ascendientes y el propio concepto.
Padres e hijos inmediatos	Capacidad para seleccionar un conjunto de conceptos, tales como los hijos y los padres inmediatos.
Conjunción	Capacidad para conectar dos restricciones de expresiones, grupos de atributos o conjuntos de atributos vía el operador lógico AND.
Disyunción	Capacidad para conectar dos restricciones de expresiones, grupos de atributos o conjuntos de atributos vía el operador lógico OR.
Refinamiento	Capacidad para refinar (i.e. especializar) el significado de una restricción de expresiones mediante uno o más valores de atributo.
Reverso	Capacidad para restringir los conceptos origen de un conjunto de relaciones y referirse a los conceptos destino de dichas relaciones.
Atributo punteado	Capacidad para referirse al valor (o conjunto de valores) de un atributo que está incluido en la definición de un conjunto de conceptos.
Grupo de atributos	Capacidad para agrupar una colección de atributos que operan conjuntamente como parte de un refinamiento.
Atributo	Capacidad para especificar un par nombre de atributo-valor de atributo que refinan el significado de las expresiones resultantes de la restricción.
Descendientes de atributo	Capacidad para definir un atributo que puede aplicar a sus descendientes o a él y a sus descendientes.

Anidamiento	Capacidad para utilizar una restricción de expresiones para representar el conjunto válido de nombres de atributo y/o valores de atributo.
Valores concretos	Capacidad para usar números enteros, decimales y cadenas como valores de atributo.
Comparador de valores de atributo	Capacidad para comparar el valor del atributo de las expresiones resultantes con el valor del atributo en la restricción de expresiones mediante operadores de comparación (i.e. =, <>, !=).
Comparador de valores concretos	Capacidad para comparar el valor del atributo de las expresiones resultantes con el valor del atributo en la restricción de expresiones mediante operadores de comparación matemáticos (i.e. =, <, >, <=, >=, !=).
Miembro de	Capacidad para seleccionar un conjunto de conceptos que están referenciados por los miembros de un conjunto de referencias (o conjunto de conjuntos de referencias).
Exclusión	Capacidad para filtrar un conjunto de expresiones del resultado mediante, o bien eliminando expresiones cuyo concepto foco está en un conjunto específico, o bien eliminando expresiones cuyo valor de atributo es igual a cierto valor.
Cualquier	Capacidad para referenciar cualquier concepto del sustrato, sin depender de la disponibilidad de un concepto raíz.

Tabla 3. Operadores y funciones requeridas en el lenguaje de restricciones

En resumen, hay tres tipos de restricciones de expresiones de SNOMED CT: simples, refinadas y compuestas. Para definir una restricción simple se aplica un operador de restricción a un concepto foco para seleccionarlo a él mismo y a sus descendientes (i.e. "<<"), solo a sus descendientes (i.e. "<"), a él y a sus ascendientes (i.e. ">>"), solo a sus ascendientes (i.e. ">") o bien seleccionar los miembros de un conjunto de referencias (i.e. "^"). Es importante recordar que también es posible seleccionar todo el sustrato de SNOMED CT con el símbolo "*" (o con "any"). Para definir restricciones refinadas es necesario aplicar un refinamiento a una restricción simple (precedido del símbolo ":"). El refinamiento está formado por uno o más pares nombre de atributo-valor de atributo. Un atributo es una relación entre conceptos de dos jerarquías y puede ir precedido por una cardinalidad que expresa el número de veces que puede repetirse dicho atributo para un mismo concepto foco (i.e. "[min..max]"), un operador de reverso (i.e. "R") o un

operador de restricción (i.e. “<<” o “<”). Hay varias opciones para el valor del atributo, a saber: una restricción simple, una restricción compuesta, una restricción refinada (i.e. anidamiento de restricciones) y un valor numérico o textual. Los pares nombre de atributo-valor de atributo se pueden combinar mediante la conjunción (i.e. “AND” o “,”) y la disyunción (i.e. “OR”) para establecer conjuntos de pares, grupos o ambos. Los grupos se procesan de un modo particular y están predefinidos en la definición lógica de los conceptos. Asimismo, los grupos pueden ir precedidos por una cardinalidad de grupo. Y, por último, las restricciones compuestas, que pueden definirse mediante conjunción, disyunción o exclusión (i.e. “MINUS”) entre restricciones simples, refinadas o ambas.

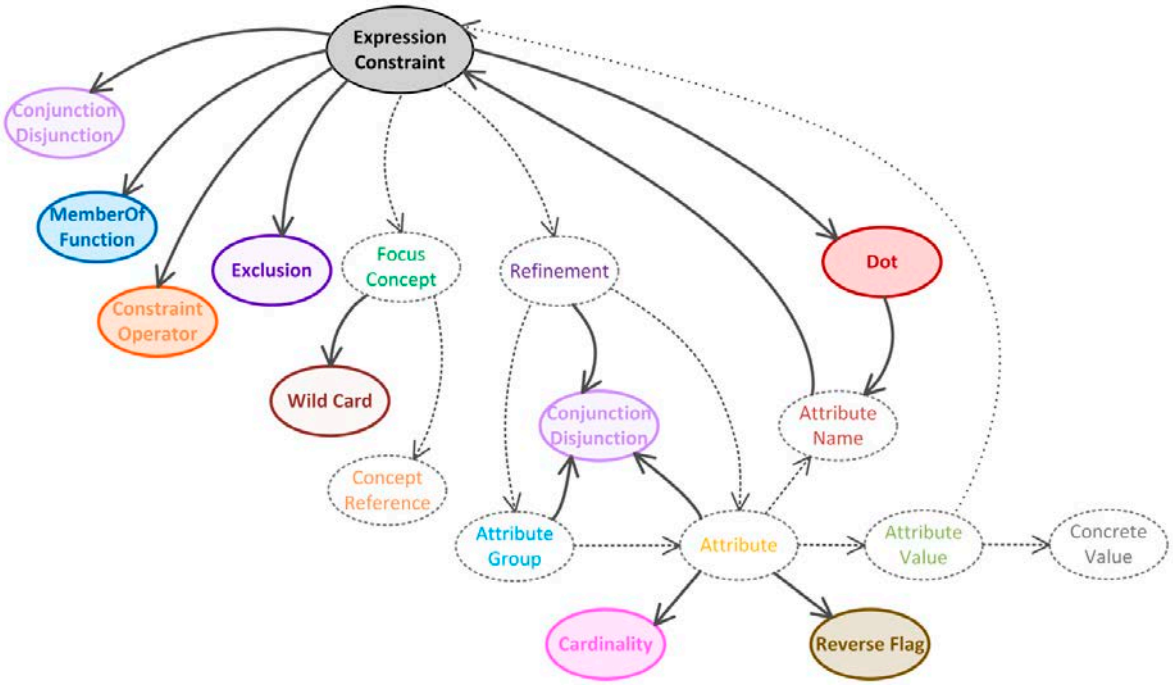

Figura 33. Modelo abstracto de una restricción de expresiones de SNOMED CT

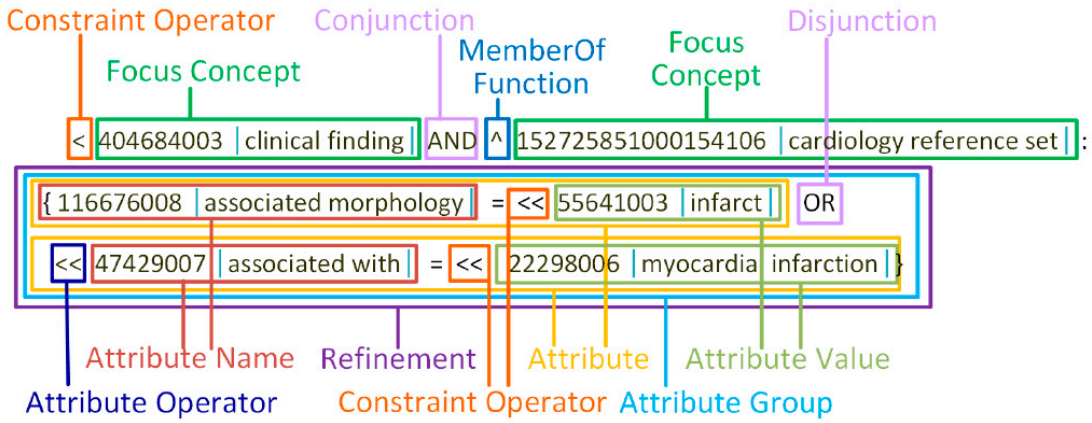

Figura 34. Componentes de una restricción de expresiones

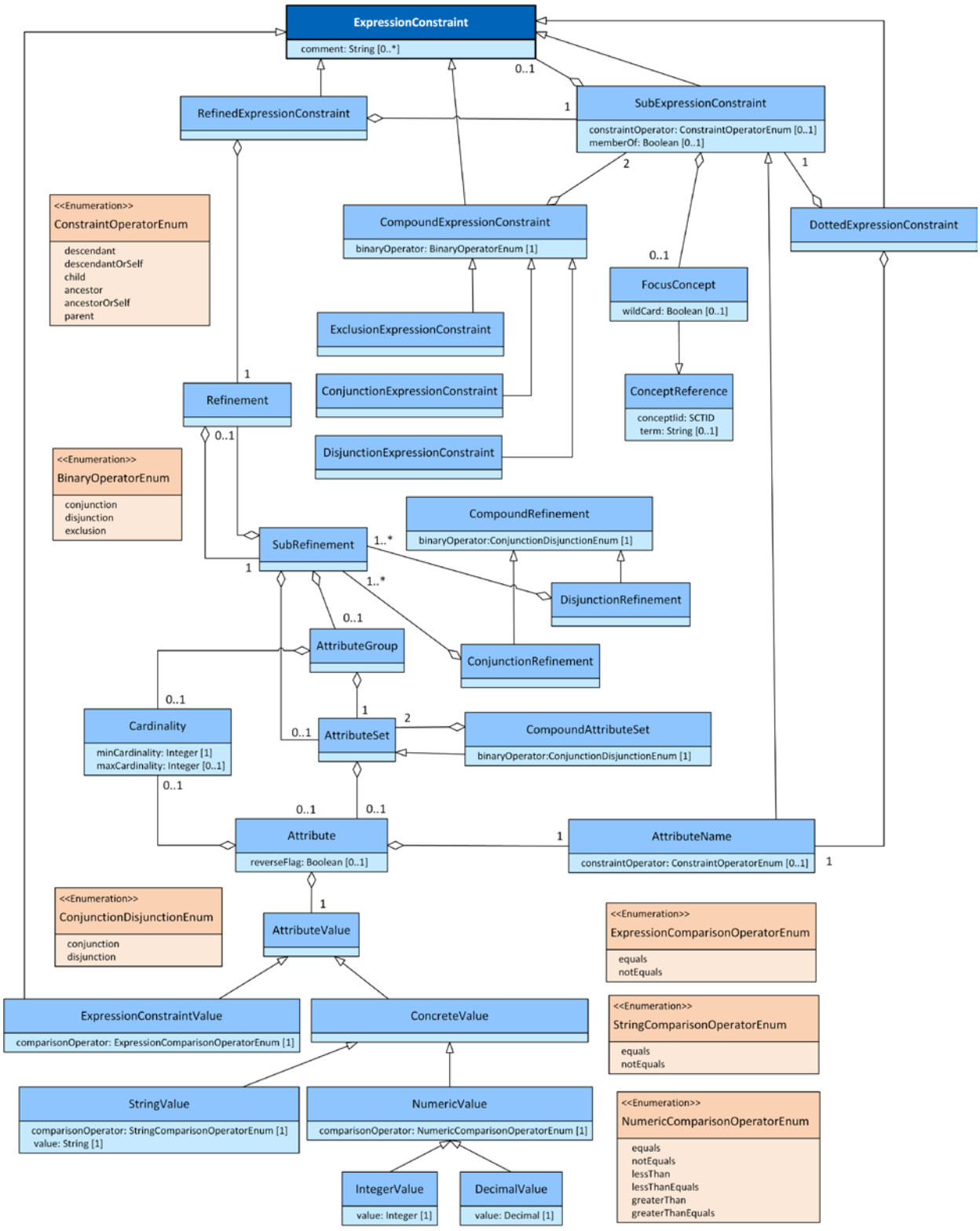

Figura 35. Modelo lógico del lenguaje de restricciones de expresiones

La siguiente definición ABNF especifica la *sintaxis corta* del lenguaje de restricciones de expresiones de SNOMED CT:

```
expressionConstraint = ws ( refinedExpressionConstraint /
compoundExpressionConstraint /
dottedExpressionConstraint / subExpressionConstraint ) ws
refinedExpressionConstraint = subExpressionConstraint ws ":" ws eclRefinement
compoundExpressionConstraint = conjunctionExpressionConstraint /
disjunctionExpressionConstraint / exclusionExpressionConstraint
conjunctionExpressionConstraint = subExpressionConstraint 1*(ws conjunction ws
subExpressionConstraint)
disjunctionExpressionConstraint = subExpressionConstraint 1*(ws disjunction ws
subExpressionConstraint)
exclusionExpressionConstraint = subExpressionConstraint ws exclusion ws
subExpressionConstraint
dottedExpressionConstraint = subExpressionConstraint 1*(ws
dottedExpressionAttribute)
dottedExpressionAttribute = dot ws eclAttributeName
subExpressionConstraint = [constraintOperator ws] [memberOf ws]
(eclFocusConcept / "(" ws
expressionConstraint ws ")")
eclFocusConcept = eclConceptReference / wildCard
dot = "."
memberOf = "^"
eclConceptReference = conceptId [ws "|" ws term ws "|"]
conceptId = sctId
term = 1*nonwsNonPipe *( 1*SP 1*nonwsNonPipe )
wildCard = "*"
constraintOperator = childOf / descendantOrSelfOf / descendantOf / parentOf /
ancestorOrSelfOf /
ancestorOf
descendantOf = "<"
descendantOrSelfOf = "<<"
childOf = "<!"
ancestorOf = ">"
ancestorOrSelfOf = ">>"
parentOf = ">!"
conjunction = (("a"/"A") ("n"/"N") ("d"/"D") mws) / ","
```

```
disjunction = ("o"/"O") ("r"/"R") mws
exclusion = ("m"/"M") ("i"/"I") ("n"/"N") ("u"/"U") ("s"/"S") mws
eclRefinement = subRefinement ws [conjunctionRefinementSet /
disjunctionRefinementSet]
conjunctionRefinementSet = 1*(ws conjunction ws subRefinement)
disjunctionRefinementSet = 1*(ws disjunction ws subRefinement)
subRefinement = eclAttributeSet / eclAttributeGroup / "(" ws eclRefinement ws ")"
eclAttributeSet = subAttributeSet ws [conjunctionAttributeSet /
disjunctionAttributeSet]
conjunctionAttributeSet = 1*(ws conjunction ws subAttributeSet)
disjunctionAttributeSet = 1*(ws disjunction ws subAttributeSet)
subAttributeSet = eclAttribute / "(" ws eclAttributeSet ws ")"
eclAttributeGroup = ["[" cardinality "]" ws] "{" ws eclAttributeSet ws "}"
eclAttribute = ["[" cardinality "]" ws] [reverseFlag ws] eclAttributeName ws
(expressionComparisonOperator ws subExpressionConstraint /
numericComparisonOperator ws "#"
numericValue / stringComparisonOperator ws QM stringValue QM)
cardinality = minValue to maxValue
minValue = nonNegativeIntegerValue
to = ".."
maxValue = nonNegativeIntegerValue / many
many = "*"
reverseFlag = "R"
eclAttributeName = subExpressionConstraint
expressionComparisonOperator = "=" / "!="
numericComparisonOperator = "=" / "!=" / "<=" / "<" / ">=" / ">"
stringComparisonOperator = "=" / "!="
numericValue = ["-"/"+"] (decimalValue / integerValue)
stringValue = 1*(anyNonEscapedChar / escapedChar)
integerValue = digitNonZero *digit / zero
decimalValue = integerValue "." 1*digit
nonNegativeIntegerValue = (digitNonZero *digit ) / zero
sctId = digitNonZero 5*17( digit )
ws = *( SP / HTAB / CR / LF / comment ) ; optional white space
mws = 1*( SP / HTAB / CR / LF / comment ) ; mandatory white space
comment = "/*" *(nonStarChar / starWithNonFSlash) "*/"
```

```
nonStarChar = SP / HTAB / CR / LF / %x21-29 / %x2B-7E /UTF8-2 / UTF8-3 /
UTF8-4
starWithNonFSlash = %x2A nonFSlash
nonFSlash = SP / HTAB / CR / LF / %x21-2E / %x30-7E /UTF8-2 / UTF8-3 / UTF8-
4
SP = %x20 ; space
HTAB = %x09 ; tab
CR = %x0D ; carriage return
LF = %x0A ; line feed
QM = %x22 ; quotation mark
BS = %x5C ; back slash
digit = %x30-39
zero = %x30
digitNonZero = %x31-39
nonwsNonPipe = %x21-7B / %x7D-7E / UTF8-2 / UTF8-3 / UTF8-4
anyNonEscapedChar = SP / HTAB / CR / LF / %x20-21 / %x23-5B / %x5D-7E /
UTF8-2 / UTF8-3 / UTF8-4
escapedChar = BS QM / BS BS
UTF8-2 = %xC2-DF UTF8-tail
UTF8-3 = %xE0 %xA0-BF UTF8-tail / %xE1-EC 2( UTF8-tail ) / %xED %x80-9F
UTF8-tail / %xEE-EF 2( UTF8-
tail )
UTF8-4 = %xF0 %x90-BF 2( UTF8-tail ) / %xF1-F3 3( UTF8-tail ) / %xF4 %x80-8F
2( UTF8-tail )
UTF8-tail = %x80-BF
```

La siguiente definición ABNF especifica la *sintaxis larga* del lenguaje de restricciones de expresiones de SNOMED CT. Conviene subrayar que la sintaxis no es sensible a mayúsculas y minúsculas (i.e. case insensitive).

```
expressionConstraint = ws ( refinedExpressionConstraint /
compoundExpressionConstraint /
dottedExpressionConstraint / subExpressionConstraint ) ws
refinedExpressionConstraint = subExpressionConstraint ws ":" ws eclRefinement
compoundExpressionConstraint = conjunctionExpressionConstraint /
disjunctionExpressionConstraint / exclusionExpressionConstraint
conjunctionExpressionConstraint = subExpressionConstraint 1*(ws conjunction ws
subExpressionConstraint)
```

```
disjunctionExpressionConstraint = subExpressionConstraint 1*(ws disjunction ws
subExpressionConstraint)
exclusionExpressionConstraint = subExpressionConstraint ws exclusion ws
subExpressionConstraint
dottedExpressionConstraint = subExpressionConstraint 1*(ws
dottedExpressionAttribute)
dottedExpressionAttribute = dot ws eclAttributeName
subExpressionConstraint = [constraintOperator ws] [memberOf ws]
(eclFocusConcept / "(" ws
expressionConstraint ws ")")
eclFocusConcept = eclConceptReference / wildCard
dot = "."
memberOf = "^" / ("m"/"M") ("e"/"E") ("m"/"M") ("b"/"B") ("e"/"E") ("r"/"R")
("o"/"O") ("f"/"F")
eclConceptReference = conceptId [ws "|" ws term ws "|"]
conceptId = sctId
term = 1*nonwsNonPipe *( 1*SP 1*nonwsNonPipe )
wildCard = "*" / ( ("a"/"A") ("n"/"N") ("y"/"Y") )
constraintOperator = childOf / descendantOrSelfOf / descendantOf / parentOf /
ancestorOrSelfOf /
ancestorOf
descendantOf = "<" / ( ("d"/"D") ("e"/"E") ("s"/"S") ("c"/"C") ("e"/"E") ("n"/"N")
("d"/"D") ("a"/"A")
("n"/"N") ("t"/"T") ("o"/"O") ("f"/"F") mws )
descendantOrSelfOf = "<<" / ( ("d"/"D") ("e"/"E") ("s"/"S") ("c"/"C") ("e"/"E")
("n"/"N") ("d"/"D")
("a"/"A") ("n"/"N") ("t"/"T") ("o"/"O") ("r"/"R") ("s"/"S") ("e"/"E") ("l"/"L") ("f"/"F")
("o"/"O") ("f"/"F")
mws )
childOf = "<!" / (("c"/"C") ("h"/"H") ("i"/"I") ("l"/"L") ("d"/"D") ("o"/"O") ("f"/"F")
mws )
ancestorOf = ">" / ( ("a"/"A") ("n"/"N") ("c"/"C") ("e"/"E") ("s"/"S") ("t"/"T")
("o"/"O") ("r"/"R") ("o"/"O")
("f"/"F") mws )
ancestorOrSelfOf = ">>" / ( ("a"/"A") ("n"/"N") ("c"/"C") ("e"/"E") ("s"/"S") ("t"/"T")
("o"/"O") ("r"/"R")
("o"/"O") ("r"/"R") ("s"/"S") ("e"/"E") ("l"/"L") ("f"/"F") ("o"/"O") ("f"/"F") mws )
```


```
parentOf = ">!" / (("p"/"P") ("a"/"A") ("r"/"R") ("e"/"E") ("n"/"N") ("t"/"T")
("o"/"O") ("f"/"F") mws )
conjunction = (("a"/"A") ("n"/"N") ("d"/"D") mws) / ","
disjunction = ("o"/"O") ("r"/"R") mws
exclusion = ("m"/"M") ("i"/"I") ("n"/"N") ("u"/"U") ("s"/"S") mws
eclRefinement = subRefinement ws [conjunctionRefinementSet /
disjunctionRefinementSet]
conjunctionRefinementSet = 1*(ws conjunction ws subRefinement)
disjunctionRefinementSet = 1*(ws disjunction ws subRefinement)
subRefinement = eclAttributeSet / eclAttributeGroup / "(" ws eclRefinement ws ")"
eclAttributeSet = subAttributeSet ws [conjunctionAttributeSet /
disjunctionAttributeSet]
conjunctionAttributeSet = 1*(ws conjunction ws subAttributeSet)
disjunctionAttributeSet = 1*(ws disjunction ws subAttributeSet)
subAttributeSet = eclAttribute / "(" ws eclAttributeSet ws ")"
eclAttributeGroup = ["[" cardinality "]" ws] "{" ws eclAttributeSet ws "}"
eclAttribute = ["[" cardinality "]" ws] [reverseFlag ws] eclAttributeName ws
(expressionComparisonOperator ws subExpressionConstraint /
numericComparisonOperator ws "#"
numericValue / stringComparisonOperator ws QM stringValue QM)
cardinality = minValue to maxValue
minValue = nonNegativeIntegerValue
to = ".." / (mws ("t"/"T") ("o"/"O") mws)
maxValue = nonNegativeIntegerValue / many
many = "*" / ( ("m"/"M") ("a"/"A") ("n"/"N") ("y"/"Y"))
reverseFlag = ( ("r"/"R") ("e"/"E") ("v"/"V") ("e"/"E") ("r"/"R") ("s"/"S") ("e"/"E")
("o"/"O") ("f"/"F")) / "R"
eclAttributeName = subExpressionConstraint
expressionComparisonOperator = "=" / "!=" / ("n"/"N") ("o"/"O") ("t"/"T") ws "=" /
"<>"
numericComparisonOperator = "=" / "!=" / ("n"/"N") ("o"/"O") ("t"/"T") ws "=" /
"<>" / "<=" / "<" / ">=" /
">"
stringComparisonOperator = "=" / "!=" / ("n"/"N") ("o"/"O") ("t"/"T") ws "=" / "<>"
numericValue = ["-"/"+"] (decimalValue / integerValue)
stringValue = 1*(anyNonEscapedChar / escapedChar)
integerValue = digitNonZero *digit / zero
```

```
decimalValue = integerValue "." 1*digit
nonNegativeIntegerValue = (digitNonZero *digit ) / zero
sctId = digitNonZero 5*17( digit )
ws = *( SP / HTAB / CR / LF / comment ) ; optional white space
mws = 1*( SP / HTAB / CR / LF / comment ) ; mandatory white space
comment = "/*" *(nonStarChar / starWithNonFSlash) "*/"
nonStarChar = SP / HTAB / CR / LF / %x21-29 / %x2B-7E /UTF8-2 / UTF8-3 /
UTF8-4
starWithNonFSlash = %x2A nonFSlash
nonFSlash = SP / HTAB / CR / LF / %x21-2E / %x30-7E /UTF8-2 / UTF8-3 / UTF8-
4
SP = %x20 ; space
HTAB = %x09 ; tab
CR = %x0D ; carriage return
LF = %x0A ; line feed
QM = %x22 ; quotation mark
BS = %x5C ; back slash
digit = %x30-39
zero = %x30
digitNonZero = %x31-39
nonwsNonPipe = %x21-7B / %x7D-7E / UTF8-2 / UTF8-3 / UTF8-4
anyNonEscapedChar = SP / HTAB / CR / LF / %x20-21 / %x23-5B / %x5D-7E /
UTF8-2 / UTF8-3 / UTF8-4
escapedChar = BS QM / BS
UTF8-2 = %xC2-DF UTF8-tail
UTF8-3 = %xE0 %xA0-BF UTF8-tail / %xE1-EC 2( UTF8-tail ) / %xED %x80-9F
UTF8-tail / %xEE-EF 2( UTF8-
tail )
UTF8-4 = %xF0 %x90-BF 2( UTF8-tail ) / %xF1-F3 3( UTF8-tail ) / %xF4 %x80-8F
2( UTF8-tail )
UTF8-tail = %x80-BF
```

Operadores de restricción

Símbolo	Significado	Ejemplo
<	Descendientes	< 69509008 |agente biológico|
<<	Él y sus descendientes	<< 69509008 |agente biológico|
>	Ascendientes	> 60482008 |aplicación de apósito en herida|
>>	Él y sus ascendientes	>> 60482008 |aplicación de apósito en herida|

Tabla 4. Operadores de restricción del lenguaje de restricciones de expresiones de SNOMED CT

Operadores conjuntistas

Símbolo	Significado	Ejemplo
AND	Intersección	<< 56208002 |úlcera| AND << 50960005 |hemorragia|
OR	Unión	<< 49872002 |virus| OR << 409822003 |bacteria|
MINUS	Diferencia	<< 19829001 |enfermedad pulmonar| MINUS << 301867009 |edema de tronco|

Tabla 5. Operadores lógicos del lenguaje de restricciones de expresiones de SNOMED CT

Ejemplos de restricciones de expresiones de SNOMED CT e interpretación gráfica

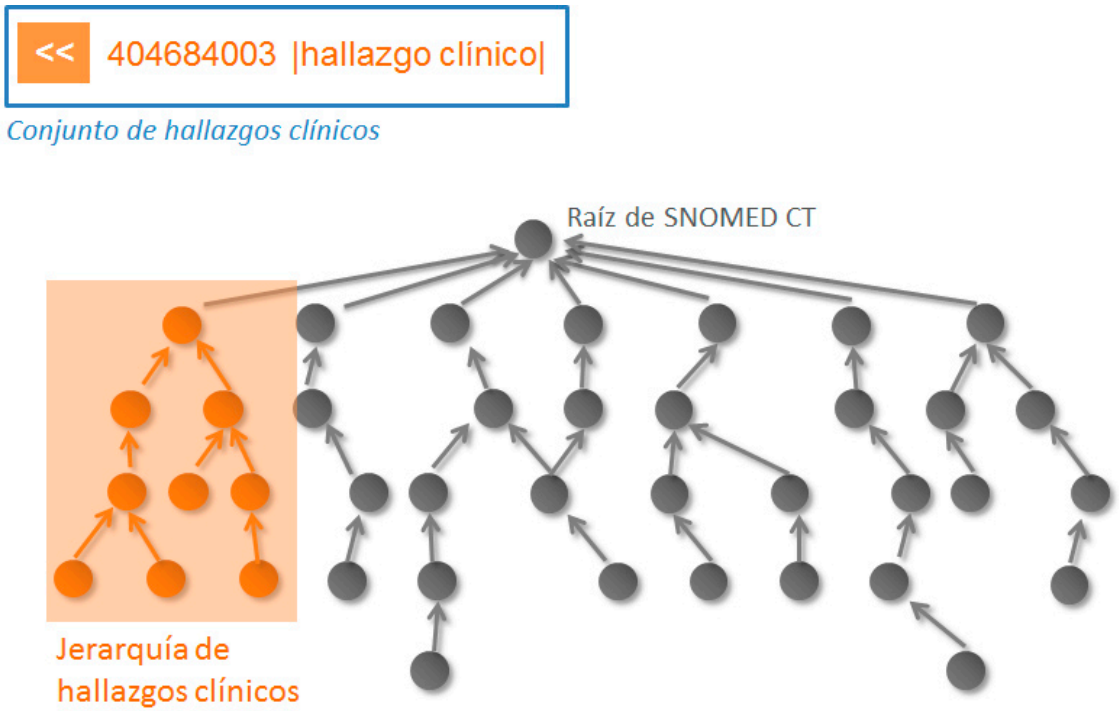

Figura 36. Conjunto de hallazgos clínicos

<< 404684003 |hallazgo clínico| OR << 71388002 |procedimiento|

Conjunto de hallazgos clínicos unido al conjunto de procedimientos

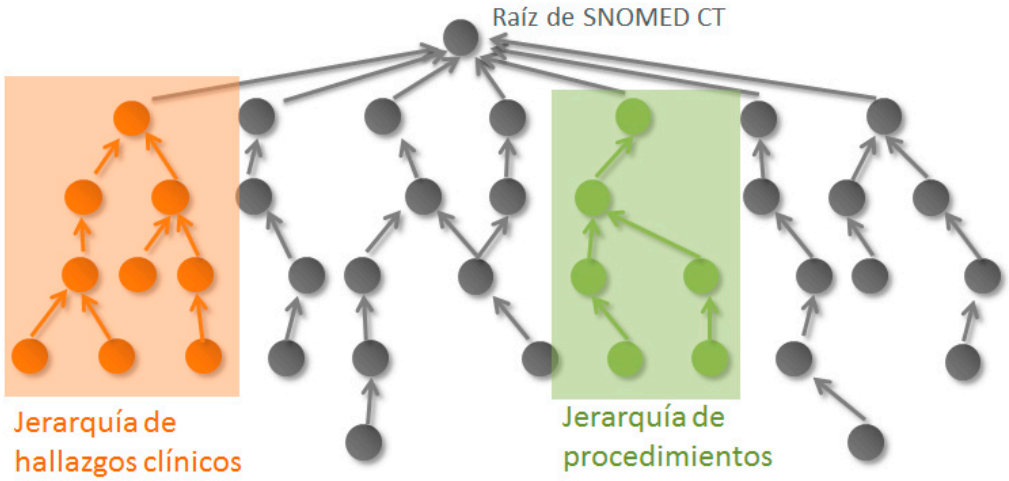

Figura 37. Conjunto de hallazgos clínicos unido al conjunto de procedimientos

< 73211009 |diabetes mellitus| :

363698007 |sitio del hallazgo| = 56329008 |glándula pituitaria|

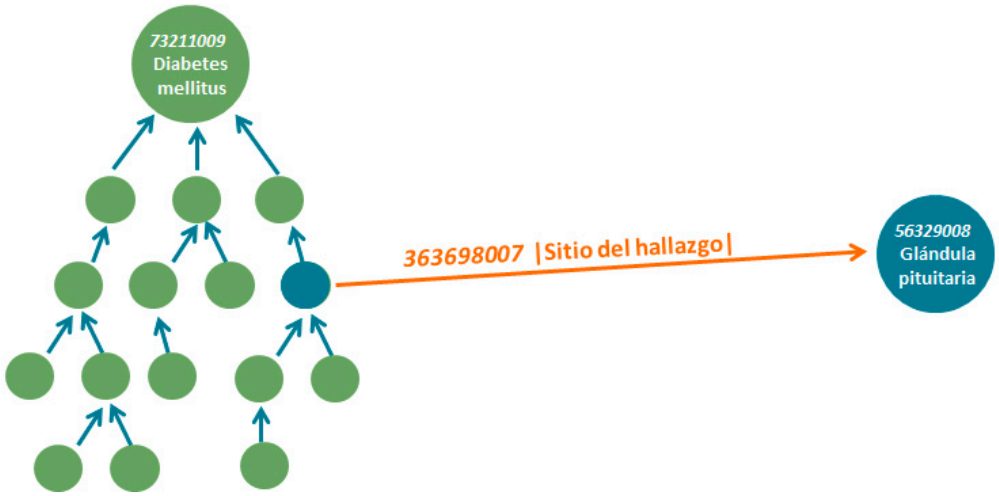

Figura 38. Conjunto de tipos de diabetes mellitus con sitio del hallazgo la glándula pituitaria

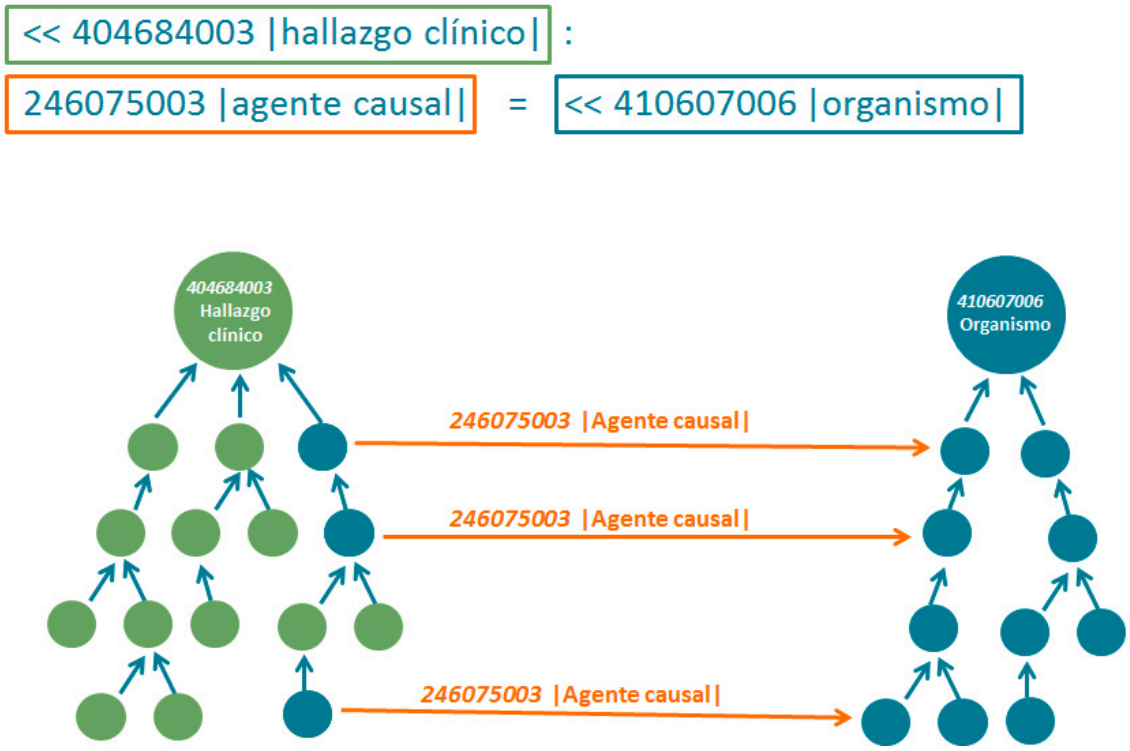

Figura 39. Conjunto de hallazgos clínicos causados por algún tipo de organismo

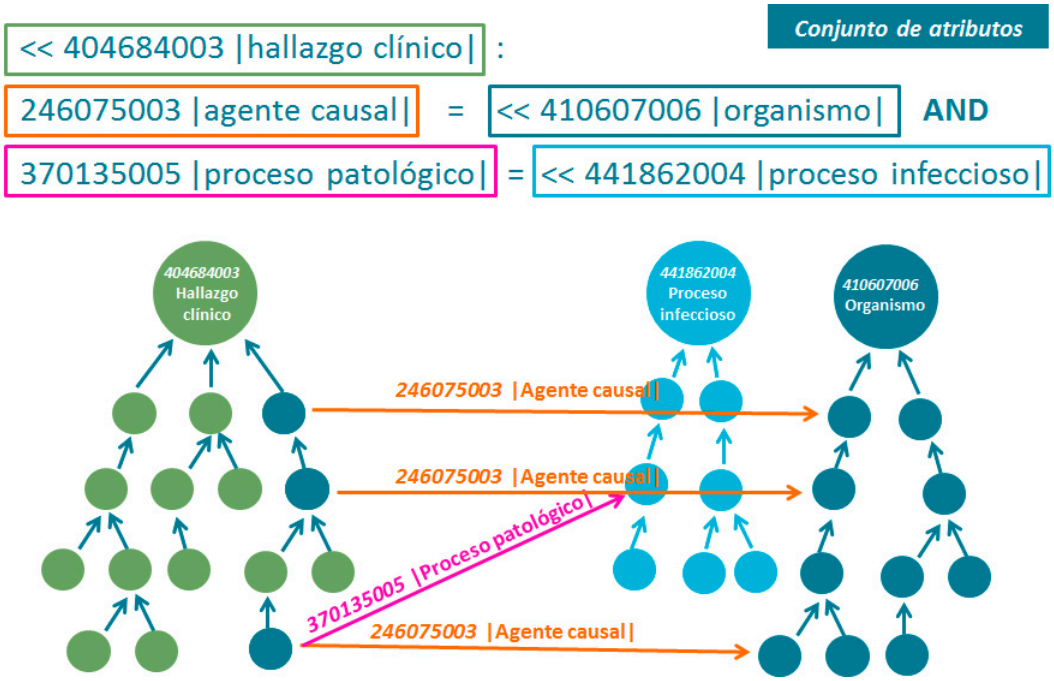

Figura 40. Conjunto de hallazgos clínicos causados por algún tipo de organismo y cuyo proceso patológico es un tipo de proceso infeccioso

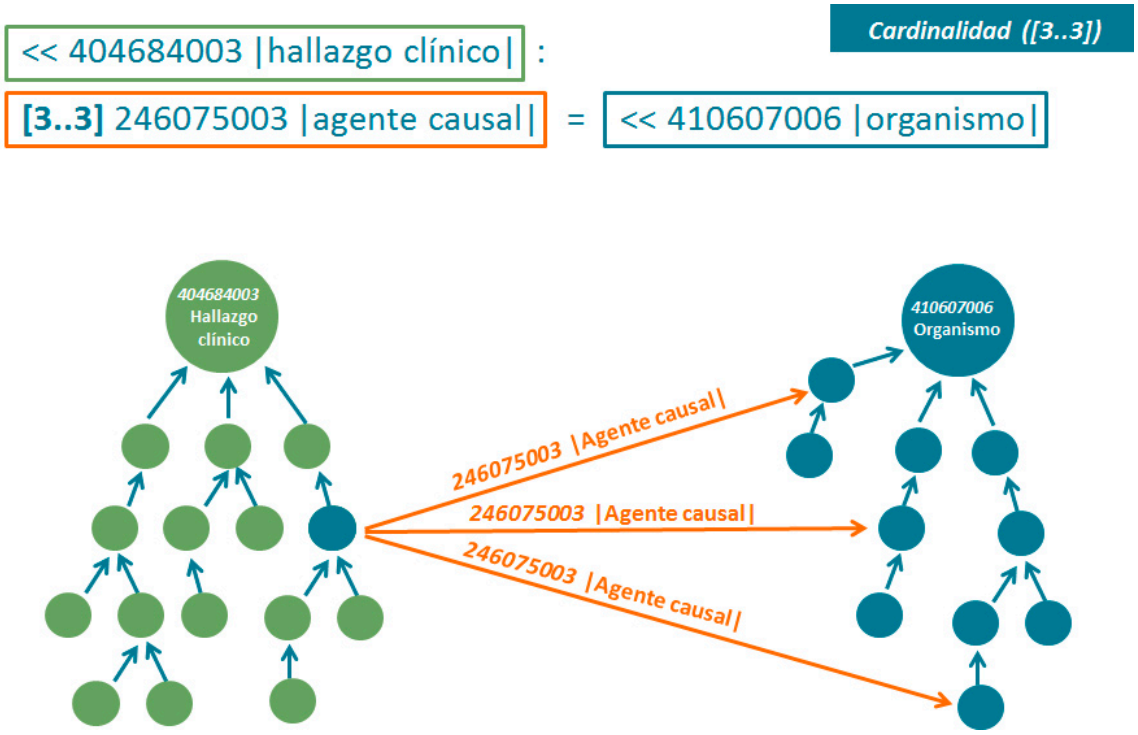

Figura 41. Conjunto de hallazgos clínicos causados exactamente por tres tipos de organismos

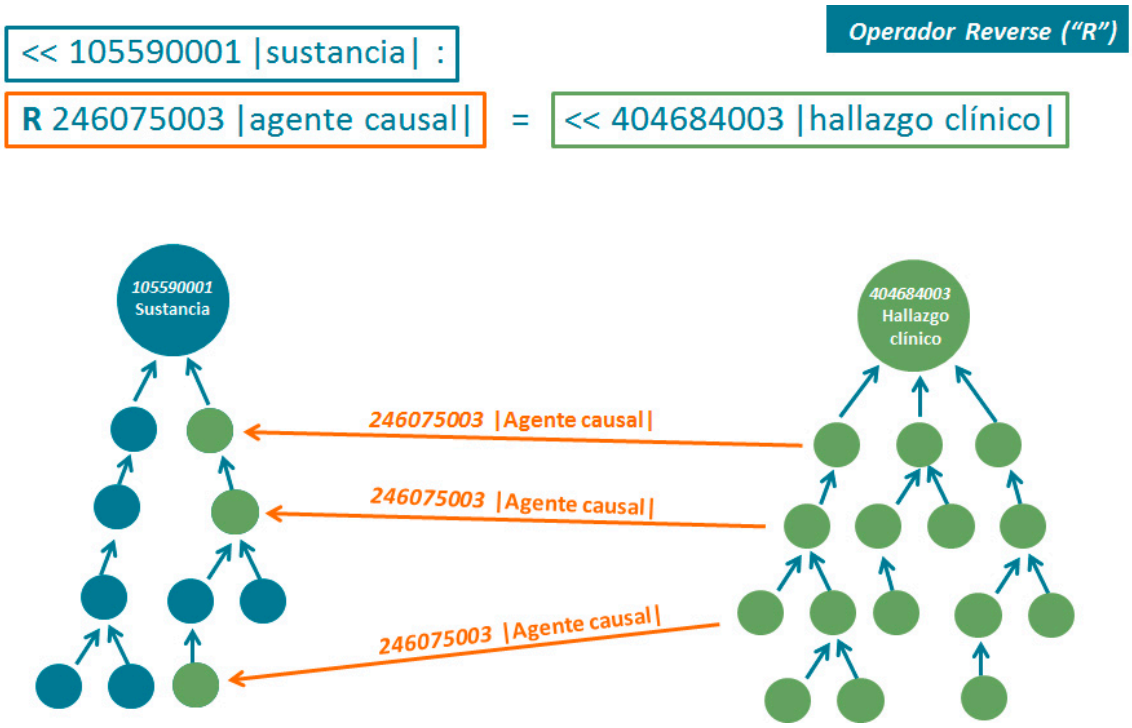

Figura 42. Conjunto de todas las sustancias que son agentes causales de hallazgos clínicos

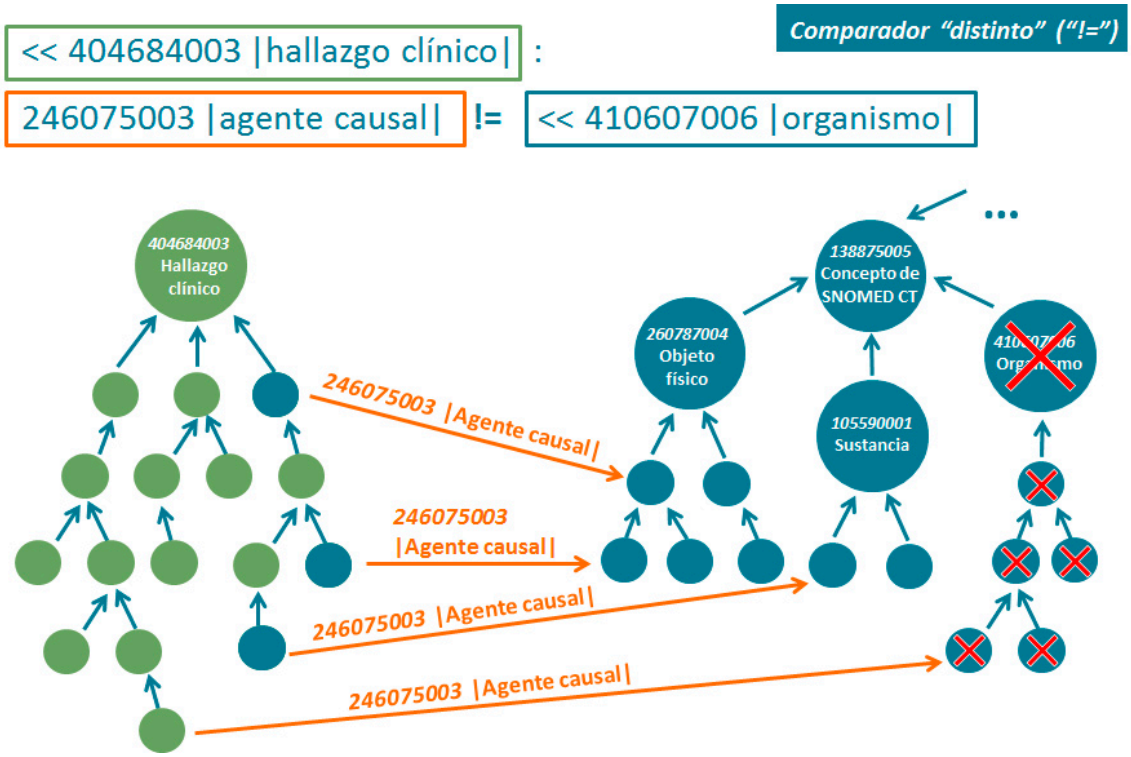

Figura 43. Conjunto de hallazgos clínicos que no están causados por ningún tipo de organismo

Más ejemplos de restricciones de expresiones de SNOMED CT

Restricción simple

< 69449002 |acción de un fármaco|

Subconjunto formado por los descendientes del concepto 'acción de un fármaco'

Restricción simple

< 247441003 |eritema|

Subconjunto formado por los descendientes del concepto 'eritema'

Restricción simple

< 8609009 |agente traumático|

Subconjunto formado por los descendientes del concepto 'agente traumático'

Restricción refinada

```
< 19829001 |enfermedad pulmonar|:
116676008 |morfología asociada| = << 79654002 |edema|
```

Subconjunto formado por las 'enfermedades de pulmón' que tienen una morfología asociada igual a 'edema' o cualquiera de sus descendientes

Restricción refinada

```
< 404684003 |hallazgo clínico|:
363698007 |sitio del hallazgo| = << 39057004 |válvula pulmonar|,
116676008 |morfología asociada| = << 415582006 |estenosis|
```

Subconjunto formado por los 'hallazgos clínicos' cuyo sitio del hallazgo es la 'estructura de la válvula pulmonar' (o un descendiente) y cuya morfología asociada es 'estenosis' (o un descendiente)

Restricción refinada

```
< 404684003 |hallazgo clínico|: 116676008 |morfología asociada| =
(< 56208002|úlcera| AND < 50960005|hemorragia|)
```

Subconjunto formado por los 'hallazgos clínicos' cuya morfología asociada está en la intersección entre los descendientes de 'úlcera' y los descendientes de 'hemorragia'

Restricción refinada

```
< 404684003 |hallazgo clínico|:
47429007 |asociado con| = (< 404684003 |hallazgo clínico|:
116676008 |morfología asociada| = < 55641003 |infarto|)
```

Subconjunto formado por los 'hallazgos clínicos' que están asociados con los 'hallazgos clínicos' cuya morfología asociada es descendiente de 'infarto'

Restricción refinada

```
< 373873005 |producto biológico/farmacéutico|:
[1..3] 127489000 |componente activo| = < 105590001 |sustancia|
```

Subconjunto formado por los 'producto biológico/farmacéutico' con uno, dos o tres sustancias como componentes activos

Restricción refinada

```
< 373873005 |producto biológico/farmacéutico|:
[2..2] 127489000 |componente activo| = < 105590001 |sustancia|
```

Subconjunto formado por los 'producto biológico/farmacéutico' con exactamente dos sustancias como componentes activos

Restricción refinada

```
<< 105590001 |sustancia|:
R 246075003 |agente causal| = < 404684003 |hallazgo clínico|
```

Subconjunto formado por las 'sustancias' que son agentes causales de los descendientes de 'hallazgo clínico'

Restricción refinada

```
< 404684003 |hallazgo clínico|:
116676008 |morfología asociada| != < 26036001 |obstrucción|
```

Subconjunto formado por los 'hallazgos clínicos' cuya morfología asociada no es un descendiente de 'obstrucción'

Restricción compuesta

```
< 19829001 |enfermedad pulmonar| AND < 301867009 |edema de tronco|
```

Subconjunto formado por la intersección entre el conjunto formado por los descendientes de 'enfermedad pulmonar' y el conjunto formado por los descendientes de 'edema de tronco'

Restricción compuesta

```
< 19829001 |enfermedad pulmonar| OR < 301867009 |edema de tronco|
```

Subconjunto formado por la unión entre el conjunto formado por los descendientes de 'enfermedad pulmonar' y el conjunto formado por los descendientes de 'edema de tronco'

Restricción compuesta

```
< 19829001 |enfermedad pulmonar| MINUS < 301867009 |edema de tronco|
```

Subconjunto formado por la diferencia entre el conjunto formado por los descendientes de 'enfermedad pulmonar' y el conjunto formado por los descendientes de 'edema de tronco'

8.2 Almacenamiento de la base de datos de SNOMED CT

Para la persistencia de SNOMED CT hemos utilizado Neo4J [11, 12], un software de base de datos orientado a grafos que incluye un potente lenguaje llamado Cypher [13] para consultar los datos almacenados en el grafo. Hemos implementado un módulo de carga Java para importar y transformar en un grafo los archivos de texto SNOMED CT que proporciona el Ministerio de Sanidad, Servicios Sociales e Igualdad español (MSSSI) y, en última instancia, SNOMED International. La última versión importada de SNOMED CT en el motor de ejecución es la INT 31/01/2017. Contiene más de 300.000 nodos (conceptos de SNOMED CT) y más de un millón y medio de relaciones entre los nodos (tanto atributos como de especialización IS A). Debido al tamaño del grafo, se ha calculado el cierre transitivo (o clausura transitiva) para agilizar el proceso de consulta (más de cinco millones de relaciones en total).

Dado que la terminología SNOMED CT está estructurada en forma de grafo –dirigido y acíclico- tiene sentido almacenarla en una base de datos orientada a grafos en lugar de un modelo de base de datos tradicional, como el modelo relacional. Las bases de datos de grafos tiene una serie de ventajas potenciales, como un mayor rendimiento al recuperar datos de una consulta, o la flexibilidad para agregar nuevos nodos y relaciones al grafo sin afectar a las consultas existentes gracias a la naturaleza aditiva de los grafos (i.e. escalabilidad) [11, 14].

Para cada nodo, además del código y su descripción de SNOMED CT (i.e. Fully Specified Name), se han añadido propiedades con su número de descendientes, la profundidad (la mínima, dado que al existir polijerarquía un nodo puede tener distintas profundidades) y la jerarquía de alto nivel a la que pertenece.

La figura siguiente muestra un fragmento del contenido de SNOMED CT. Las relaciones IS A punteadas indican que existen nodos en ese camino pero no están representados para simplificar la figura.

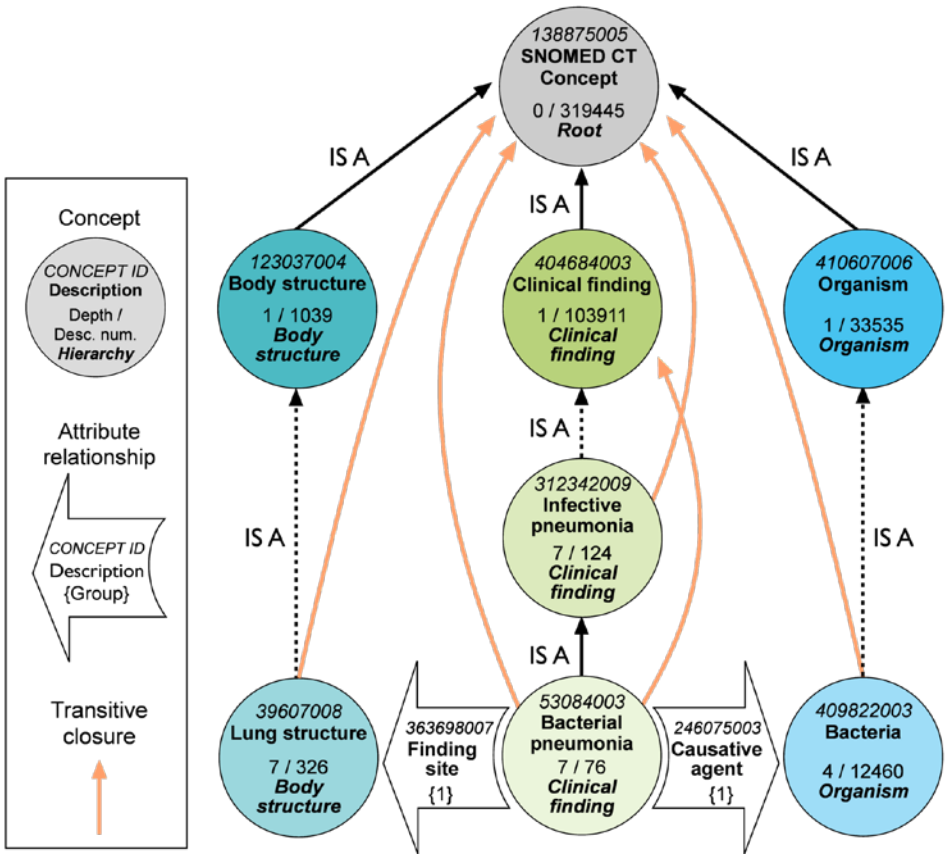

Figura 44. Ejemplo de un subconjunto de nodos de SNOMED CT con sus propiedades y sus relaciones jerárquicas (incluyendo el cierre transitivo) y de atributo. Cada nodo contiene su código, su descripción completa, su profundidad (mínima) en el grafo, su número de descendientes y la jerarquía a la que pertenece. Cada atributo contiene su código, su descripción completa y el número de grupo

9. Resultados

9.1 Implementación de un módulo Java de importación de SNOMED CT

El módulo de importación está formado por dos clases Java: una para las ediciones internacional y española, y otra para la extensión del Sistema Nacional de Salud español (SNS). Se ha optado por separar el proceso de importación en dos clases, dado que el tratamiento de los ficheros con el sustrato de SNOMED CT difiere según se trate de una edición o una extensión, si bien es cierto que ambos procesos tienen partes comunes.

Los ficheros de entrada del módulo se obtienen del área de descarga de SNOMED CT del sitio web del Ministerio de Sanidad, Servicios Sociales e Igualdad español (MSSSI)[1] . El formato de los ficheros es 'Release Format 2' (RF2)[2] en su versión 'Snapshot'[3] . Los ficheros base necesarios para generar la edición internacional son tres: conceptos, descripciones y relaciones. La edición española tiene su propio fichero de descripciones y aprovecha los ficheros de conceptos y de relaciones de la edición internacional. A su vez, la extensión del SNS tiene tres ficheros con conceptos, descripciones y relaciones, respectivamente. Estos ficheros suponen una extensión al núcleo de SNOMED CT, por lo que es necesario llevar a cabo una unión entre los ficheros de conceptos y de relaciones, y los homólogos de la edición internacional. Dado que se producen algunas disimilitudes de vocabulario entre el español de América Latina y el español de España, el fichero de descripciones del SNS se suma al de la edición española teniendo en cuenta que un subconjunto de sus descripciones sobrescriben a las descripciones de la edición española (e.g. 102261002 |Frutilla| de la edición española pasa a ser 102261002 |Fresa| en la extensión del SNS; o 256349002 |Maní|, que pasa a ser 256349002 |Cacahuete|, entre otros).

El siguiente paso es la generación de sendos ficheros con la clausura transitiva de los conceptos; uno para las ediciones internacional y española, y otro para la extensión del SNS. Para ello, se hace uso de un script de Perl, disponible en el apartado de recursos técnicos de la web de SNOMED CT Internacional[4] . Dicho script toma como entrada los ficheros de relaciones de las ediciones internacional y SNS (esta última, ya unida a la edición internacional), y genera los ficheros con las respectivas clausuras transitivas. Además, el script ha sido modificado para añadir el número de descendientes de cada

[1] https://www.msssi.gob.es/profesionales/hcdsns/areaRecursosSem/snomed-ct/areaDescarga.htm

[2] IHTSDO propuso pasar a estado 'deprecated' el formato RF1 en junio de 2015: http://www.snomed.org/news-articles/proposed-deprecation-and-withdrawal-of-support-of-snomed-ct-release-format-1-rf1

[3] IHTSDO proporciona además versiones 'Delta' y 'Full'. Más información sobre los ficheros en https://confluence.ihtsdotools.org/display/DOCRELFMT/2.1.+SNOMED+CT+-+File+Naming+Conventions

[4] https://confluence.ihtsdotools.org/display/DOC/Technical+Resources

concepto, dado que, en nuestro caso particular, será una propiedad más de los conceptos en la base de datos.

A continuación, se genera un fichero con la profundidad mínima de cada concepto[5] y otro con la jerarquía de alto nivel a la que pertenece cada concepto, ya que, como en el caso anterior, serán propiedades que añadiremos a los conceptos en la base de datos. Este proceso, como antes, se lleva a cabo tanto para las ediciones internacional y española, como para la extensión del SNS.

Una vez finalizado el proceso automático de pre-procesado y creación de ficheros auxiliares por parte del módulo, solo resta generar la base datos a partir de ellos. Dada la estructura interna de la terminología SNOMED CT (i.e. grafo dirigido y acíclico), se ha optado por generar una base de datos orientada a grafos. En nuestro caso particular, se ha utilizado Neo4j, que es un software libre de bases de datos orientadas a grafos que ofrece su propio lenguaje de consultas (i.e. Cypher)[6] . Para ello, las clases Java del módulo, una vez generada la base de datos con el grafo vacío y los índices programados, ejecutan una serie de consultas Cypher con el siguiente orden y propósito: creación de los conceptos en forma de nodos del grafo; adición de las descripciones[7] ; adición de las relaciones de atributo; adición de las relaciones de jerarquía (i.e. IS A); adición de las relaciones de clausura transitiva; adición de una propiedad a cada nodo con su número de descendientes; adición de una propiedad a cada nodo con la jerarquía de alto nivel a la que pertenece; y, por último, adición de una propiedad a cada nodo con su profundidad mínima dentro del grafo.

9.1.1 Tiempos medios de creación de la base de datos

Los tiempos medios de creación de la base de datos de SNOMED CT, desglosados por pasos, se recogen a continuación para una máquina con las siguientes características:

Sistema operativo: Windows 10 Education
Procesador: Intel Core i7-6700HQ CPU @ 2.60 GHz
RAM: 16 GB
Disco duro: 1 T
Tipo de sistema: 64 bits

[5] Se habla de profundidad 'mínima' puesto que SNOMED CT es un grafo que ofrece polijerarquías. Es decir, un nodo puede tener varios padres. Por lo tanto, un nodo puede tener varias profundidades.

[6] https://neo4j.com/

[7] Fully Specified Name (FSN)

Paso	Tiempo (en segundos)[8]
Creación del grafo vacío	13.198
Creación de índices en aristas	0.586
Creación de índices en nodos	0.005
Creación de nodos	88.63
Asignación de la descripción de cada nodo	27.541
Creación de relaciones de atributo	103.562
Creación de relaciones de jerarquía IS A	108.403
Creación del cierre transitivo	1693.782
Creación de propiedad con el número de descendientes en cada nodo	18.991
Creación de propiedad con la jerarquía de alto nivel a la que pertenece cada nodo	35.442
Creación de propiedad con la profundidad mínima de cada nodo	390.374
TOTAL	**2480.514**

Tabla 6. Pasos en la creación de la base de datos de SNOMED CT y tiempos medios en segundos

9.2 Desarrollo de un motor de ejecución de restricciones de expresiones de SNOMED CT

9.2.1 Generalidades

Cuando se implementa el lenguaje de restricciones de expresiones de SNOMED CT, los factores que hay que tener en consideración dependen del tipo de tareas que se quieren llevar a cabo. Por ejemplo, las implementaciones pueden requerir que las restricciones de expresiones sean creadas, parseadas sintácticamente, validadas semánticamente, ejecutadas, mostradas o intercambiadas.

[8] El punto hace referencia a los decimales

Para poder definir subconjuntos de conceptos de SNOMED CT mediante el lenguaje de restricciones de expresiones de SNOMED CT, ha sido necesario desarrollar un motor de ejecución [15] que dé soporte a este lenguaje. Gracias a estos subconjuntos de conceptos va a ser posible definir enlaces terminológicos de contenido entre modelos de información clínicos, tales como arquetipos, y la terminología SNOMED CT. Dichos modelos junto al enlace de contenido constituyen los pilares necesarios para poder alcanzar un nivel alto en interoperabilidad semántica.

El motor de ejecución permite parsear sintácticamente y ejecutar las restricciones de expresiones. Es importante destacar que el motor está en su versión beta y soporta aproximadamente el 75% de la sintaxis del lenguaje, dado que SNOMED International ha ido ampliando los operadores y la funcionalidad del lenguaje a medida que ha ido lanzando versiones. Las funcionalidades que no soporta son: cardinalidad de grupos, operador de restricción de atributo, atributo punteado, operador "miembros de" y valores numéricos y textuales como valores de atributos. Asimismo, hay algunos tipos de restricciones de expresiones, especialmente aquellas que contienen cardinalidades, cuya estrategia de ejecución está bajo revisión. La siguiente tabla recoge las partes de la sintaxis a las que se da soporte en el motor de ejecución y las que está previsto abordar como trabajo futuro.

Función	**Detalles**	**Soportado**
Referencia a concepto	Capacidad para referenciar conceptos pre-coordinados de SNOMED CT usando su identificador y una descripción opcional en lenguaje natural.	
Jerarquías de conceptos	Capacidad para seleccionar un conjunto de conceptos, como son los descendientes de un concepto, los descendientes y el propio concepto, los ascendientes y los ascendientes y el propio concepto.	
Padres e hijos inmediatos	Capacidad para seleccionar un conjunto de conceptos, tales como los hijos y los padres inmediatos.	
Conjunción	Capacidad para conectar dos restricciones de expresiones, grupos de atributos o conjuntos de atributos vía el operador lógico AND.	
Disyunción	Capacidad para conectar dos restricciones de expresiones, grupos de atributos o conjuntos de atributos vía el operador lógico OR.	

Refinamiento	Capacidad para refinar (i.e. especializar) el significado de una restricción de expresiones mediante uno o más valores de atributo.	
Reverso	Capacidad para restringir los conceptos origen de un conjunto de relaciones y referirse a los conceptos destino de dichas relaciones.	
Atributo punteado	Capacidad para referirse al valor (o conjunto de valores) de un atributo que está incluido en la definición de un conjunto de conceptos.	
Grupo de atributos	Capacidad para agrupar una colección de atributos que operan conjuntamente como parte de un refinamiento.	
Atributo	Capacidad para especificar un par nombre de atributo-valor de atributo que refinan el significado de las expresiones resultantes de la restricción.	
Descendientes de atributo	Capacidad para definir un atributo que puede aplicar a sus descendientes o a él y a sus descendientes.	
Anidamiento	Capacidad para utilizar una restricción de expresiones para representar el conjunto válido de nombres de atributo y/o valores de atributo.	
Valores concretos	Capacidad para usar números enteros, decimales y cadenas como valores de atributo.	
Comparador de valores de atributo	Operador de comparación entre atributo y rango (cuando el rango son **conceptos**) (i.e. =, <>, !=).	
Comparador de valores concretos	Operador de comparación entre atributo y rango (cuando el rango son **números o cadenas de texto**) (i.e. =, <, >, <=, >=, !=).	

Miembro de	Capacidad para seleccionar un conjunto de conceptos que están referenciados por los miembros de un conjunto de referencias (o conjunto de conjuntos de referencias).	
Exclusión	Capacidad para filtrar un conjunto de expresiones del resultado mediante, o bien eliminando expresiones cuyo concepto foco está en un conjunto específico, o bien eliminando expresiones cuyo valor de atributo es igual a cierto valor.	
Cualquier	Capacidad para referenciar cualquier concepto del sustrato, sin depender de la disponibilidad de un concepto raíz.	
Cardinalidad de atributos	Capacidad para poder definir el número máximo, mínimo o un valor concreto en el número de veces que puede repetirse una relación de atributo para un concepto.	

Tabla 7. Funciones soportadas y no soportadas en el motor de ejecución del lenguaje de restricciones de expresiones de SNOMED CT

El motor de ejecución de restricciones de expresiones de SNOMED CT está implementado en Java (se ha utilizado la plataforma Eclipse) y está disponible vía interfaz web en http://diebosto2.pc.upv.es:8888/SnomedQuery/ para su uso libre. La interfaz ha sido desarrollada con el framework para aplicaciones web open-source para el desarrollo de aplicaciones web, Vaadin. El presente trabajo se centra únicamente en la implementación del motor. La parte correspondiente al desarrollo de la interfaz (así como un servicio web disponible para realizar llamadas al motor) ha sido llevada a cabo con el apoyo de miembros del Grupo de Informática Biomédica (IBIME) del instituto ITACA de la Universitat Politècnica de València, donde ha sido llevado a cabo este trabajo.

Las descripciones de los conceptos en las restricciones de expresiones son opcionales, tal y como marca la sintaxis del lenguaje, y el motor no las tiene en cuenta para evaluar (i.e. ejecutar) las restricciones. El motor, tras ejecutar la restricción, retorna la lista de los códigos de los conceptos que conforman el subconjunto definido por la restricción, junto a su descripción completa (i.e. Fully Specified Name). La siguiente figura muestra los resultados tras ejecutar la restricción de expresiones < 19829001 |disorder of lung| AND < 301867009 |edema of trunk| en el motor de ejecución.

Figura 45. Interfaz del motor de ejecución de restricciones de expresiones de SNOMED CT tras ejecutar < 19829001 |disorder of lung| AND < 301867009 |edema of trunk|

9.2.2 Validación de las restricciones de expresiones

Dos tipos de validación son requeridas previas a la ejecución de una restricción de expresiones de SNOMED CT.

La primera de ellas, y de carácter obligatorio, es la *validación sintáctica*. Para ello se ha implementado un parseador con JavaCC (Java Compiler Compiler) [16], un generador de código abierto de parseadores para utilizar en aplicaciones desarrolladas en Java. El parseo es el proceso de analizar una cadena de caracteres consecuentemente a las reglas de una gramática formal. El proceso de parsear una restricción de expresiones de SNOMED CT implica procesar la cadena con la restricción de expresiones utilizando una de las especificaciones de la sintaxis en forma ABNF que se ha tratado con anterioridad y dividiéndola en sus partes constituyentes. Esto crea una representación de la restricción de expresiones que se puede procesar posteriormente. El análisis de una restricción de expresiones se requiere para realizar la validación sintáctica, validación semántica o ejecución de modelos conceptuales. Debe tenerse en cuenta, al analizar, que el lenguaje no distingue entre mayúsculas y minúsculas. Este proceso crea un árbol de objetos que se almacena en memoria con la restricción de expresiones para procesarlo posteriormente.

En segundo lugar, la *validación semántica*, es opcional (aunque recomendable). Tal y como hemos visto antes, el modelo de conceptos de SNOMED CT procesable por ordenador (Machine Readable Concept Model - MRCM) establece las reglas semánticas procesables por ordenador requeridas para realizar la validación semántica de SNOMED

CT, tanto en post-coordinaciones como en restricciones de expresiones (por ejemplo, el atributo 'sitio del hallazgo' de un 'hallazgo clínico' no puede ser una 'sustancia'). El MRCM está en desarrollo en este momento por parte de SNOMED International. Como consecuencia, la validación semántica no estará disponible hasta que un MRCM estable esté listo.

9.2.3 Traducción al lenguaje Cypher de Neo4j

Las restricciones de expresiones de se ejecutan sobre la base de datos de SNOMED CT explicada anteriormente para obtener los conceptos resultantes. Para llevar a cabo esta ejecución, y dado que la base de datos está almacenada en una base de datos orientada a grafos de Neo4j, se realiza un proceso de traducción del lenguaje de restricciones al lenguaje Cypher de consultas de bases de datos Neo4j. Las consultas Cypher se ejecutan sobre la base de datos de grafo de Neo4j, es decir, sobre el sustrato de SNOMED CT y como resultado se obtiene el subconjunto de conceptos definido intensionalmente (i.e. por comprensión) definido por la restricción de expresiones. Los conceptos se muestran en la interfaz del motor de ejecución en forma tabular.

9.2.4 Fases de ejecución

La ejecución de una restricción de SNOMED CT se divide en dos fases. La primera de ellas es la obtención de los identificadores numéricos, y es, por tanto, obligatoria. La segunda es la obtención de las descripciones. Esta fase es opcional y depende de las necesidades del usuario. Por defecto, la interfaz web del motor obtiene tanto los identificadores numéricos como las descripciones.

El motor de ejecución no tiene en cuenta la descripción de los conceptos. Esto significa que los usuarios pueden introducir las restricciones de expresiones con o sin descripción e incluso con una descripción incorrecta. Al final de la primera fase del proceso de ejecución, el motor devuelve la lista de identificadores numéricos que representan el subconjunto de conceptos. Sin embargo, es posible obtener la descripción completa (i.e. Fully Specified Name) de cada concepto del subconjunto en una segunda fase. Para ello, el motor vuelve a consultar el grafo, pero esta vez con el subconjunto de identificadores numéricos obtenidos en la primera fase como parámetro. Esta segunda fase tiene un coste temporal lineal con el número de conceptos en el subconjunto obtenido en la fase 1. Debido a este coste, que es significativo para subconjuntos grandes, se ha añadido al motor la posibilidad de combinar estas dos fases en sólo una fase que devuelve tanto los identificadores de conceptos numéricos como las descripciones completas de cada concepto del subconjunto cuando las restricciones de expresiones de SNOMED CT consisten en una jerarquía completa o una combinación de algunas jerarquías, sin refinamientos (es decir, una restricción simple o una compuesta sin ningún refinamiento), como por ejemplo:

1) * (retorna todo el sustrato de SNOMED CT)
2) < 404684003 |Clinical finding| (retorna toda la jerarquía de hallazgos clínicos)
3) < 19829001 |Disorder of lung| OR < 301867009 |Edema of trunk| (retorna la unión de las jerarquías enfermedad de pulmón y edema de tronco)

9.2.5 Optimizaciones

Debido a la potencia del lenguaje de restricciones de expresiones de SNOMED CT y al tamaño de la terminología de SNOMED CT (más de 300000 conceptos), es posible construir restricciones de expresiones que involucren grandes jerarquías, incluso incluyendo todo SNOMED CT. Por ejemplo, la restricción refinada siguiente especifica el subconjunto SNOMED CT de todos los conceptos en la terminología que son causados por cualquier subtipo de bacterias:

<138875005 |SNOMED CT Concept|:

246075003 |Causative agent| = <409822003 |Superkingdom Bacteria|

El modelo de conceptos de SNOMED CT es el conjunto de reglas que rigen la semántica de la terminología. Especifica los conjuntos de atributos permitidos que se pueden aplicar a cada jerarquía (dominio) y las jerarquías permitidas para cada atributo (rango). Teniendo en cuenta el modelo de conceptos de SNOMED CT (debe tenerse en cuenta que, para el propósito de la optimización, el motor de ejecución utiliza el modelo de conceptos de SNOMED CT que se presenta en guía técnica de implementación de SNOMED CT, y no la versión MRCM que saldrá en breve por parte de SNOMED International), es posible optimizar la restricción de expresión anterior. De acuerdo con sus reglas, el atributo "246075003 | Agente causante |" sólo es aplicable a las jerarquías de alto nivel de hallazgo clínico y eventos. Así, internamente, el motor sólo tiene en cuenta estas dos jerarquías (hallazgo clínico y evento) como el dominio de la restricción de expresión para calcular los conceptos resultantes en lugar de recorrer todas las jerarquías de la terminología. El mismo proceso se puede hacer en el lado del rango. Por ejemplo, la siguiente restricción de expresión refinada define el subconjunto de todas las estructuras anatómicas con lateralidad:

< 91723000 |Anatomical structure|:

272741003 |Laterality| = <138875005 |SNOMED CT Concept|

De nuevo, es posible seguir las reglas que especifica el modelo de conceptos de SNOMED CT y simplificar la restricción de expresión refinada usando el rango permitido para el atributo "272741003 | Laterality |", que es la jerarquía "<< 182353008 | Side | en lugar de usar todo el sustrato de SNOMED CT como el rango de la restricción de expresión.

9.2.6 Interfaz de usuario

Para ejecutar una restricción de SNOMED CT solo es necesario escribir una restricción de expresión en el cuadro de texto de entrada y ejecutar el motor pulsando el botón Ejecutar. También es posible elegir entre la validación sintáctica y la evaluación (i.e. ejecución) de la restricción de expresión o solo la validación sintáctica (hay que tener en cuenta que es necesario evaluar la restricción de expresión para obtener el subconjunto de conceptos correspondiente).

Edición de las restricciones de expresiones

La hándicap más importante al editar restricciones de expresiones de SNOMED CT con el motor de ejecución es el conocimiento obligatorio del lenguaje de restricciones de expresiones de SNOMED CT. Debido a esto, actualmente se están estudiando algunas posibilidades a la hora de editar restricciones de expresión de SNOMED CT de una manera guiada y amigable. Una de ellos consiste en una representación gráfica de árbol binario de restricciones de expresiones, donde los operadores, atributos y conceptos se combinan de acuerdo con la sintaxis del lenguaje para formar bloques o plantillas que se utilizan comúnmente en las restricciones de expresiones, o incluso se pueden utilizar como elementos atómicos para construir restricciones de expresiones más complejas. Estos bloques se pueden combinar con conceptos, atributos y operadores por medio de unos pocos clics del ratón, para componer la restricción de expresión requerida. Gracias a estos bloques o plantillas no es estrictamente necesario tener un profundo conocimiento del idioma (aunque un conocimiento básico es aconsejable).

Resultados

Los conceptos resultantes (es decir, el subconjunto de conceptos de SNOMED CT definidos por las restricciones de expresiones) se muestran en la interfaz en forma tabular. Además del identificador numérico del concepto y la descripción completa (i.e. Fully Specified Name), cada concepto está vinculado al navegador de SNOMED CT de SNOMED CT International para explorar sus detalles, a saber: un resumen del concepto (incluyendo sus relaciones con otros conceptos), sinónimos y aceptabilidad, diagrama de su definición lógica formal y alguna otra información útil. Además, la interfaz del motor de ejecución ofrece la posibilidad de exportar los subconjuntos a formatos de archivo TXT, XLS y CSV (es importante leer atentamente la licencia del motor de ejecución antes de exportar subconjuntos en la medida en que es un sistema que incluye una base de datos SNOMED CT).

9.2.7 Visualización de los grafos resultantes

Además de la representación tabular de los subconjuntos SNOMED CT, se ha implementado e incluido en la interfaz del motor de ejecución tres visualizaciones gráfica de subconjuntos. De momento las tres son experimentales. Conviene subrayar que dichas representaciones se han llevado a cabo con el apoyo de miembros del Grupo de Informática Biomédica (IBIME) del instituto ITACA de la Universitat Politècnica de València, donde ha sido llevado a cabo este trabajo.

Para la primera representación se ha utilizado una versión modificada del plug-in de visualización de proyectos de InteractiveVis de Oxford Internet Institute y la biblioteca de Sigma JavaScript que se dedica a dibujar grafos y publicar redes en páginas web y también para integrar exploraciones de redes en aplicaciones web enriquecidas. También se ha incluido la opción de invocar ForceAtlas2, utilizado para la distribución espacial de la red de la biblioteca Sigma.

El plug-in del proyecto InteractiveVis permite la visualización de nodos y relaciones, buscar un nodo especificado, jerarquías de grupo usando colores y otras opciones útiles como el zoom. La siguiente figura muestra una visualización del subconjunto definido por los descendientes de la diabetes mellitus y de sí mismo (<< 73211009 | diabetes mellitus |).

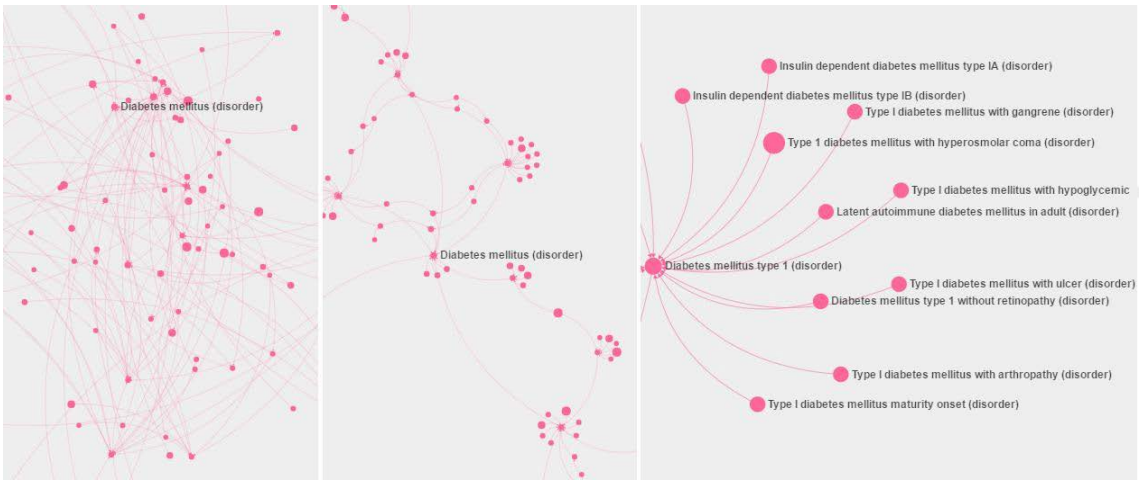

Figura 46. Visualización gráfica de << 73211009 | diabetes mellitus |, sin utilizar ForceAtlas2 (izquierda) y utilizándolo (centro). Una vista detallada parcial de la sub-jerarquía diabetes mellitus tipo I se muestra en el lado derecho de la figura

La siguiente visualización gráfica se ha llevado a cabo con otra librería de dibujado de grafos (i.e. Cinnamon). Con esta representación solo interesa mostrar los nodos intermedios del grafo (es decir, todos los nodos a excepción de los nodos hoja) en aras de intentar situar gráficamente en qué lugar de SNOMED CT se halla ubicado el subconjunto de conceptos resultado de ejecutar la restricción de expresiones de SNOMED CT, con la particularidad que solo se muestran aquellos nodos intermedios que, o bien pertenecen al subconjunto resultado, o bien tiene uno o más descendientes que pertenecen. Además, se muestra información extra, lo que supone un valor añadido

a la representación anterior. Concretamente, para cada nodo (que aparece como una burbuja susceptible de ser explotada, en su caso) se calcula y se muestra qué porcentaje de sus descendientes pertenecen al subconjunto resultado, y del subconjunto resultado, qué porcentaje pertenece a los descendientes de dicho nodo. Con esto podemos establecer los nodos más relevantes de la representación gráfica mostrada. Asimismo, es posible mostrar los nodos más relevantes (hasta cinco niveles), es posible simplificar, en su caso, la restricción de expresiones una vez se tiene calculado el subconjunto resultado (i.e. simplificación post-ejecución) y mostrar los nodos con más polijerarquía. Esta herramienta permite incluso, potencialmente, encontrar anomalías o conceptos "outliers" en la terminología SNOMED CT. De modo que podría incluso llevar a cabo evaluaciones sobre la consistencia de la propia terminología. La siguiente figura muestra la representación del subconjunto definido por la restricción de expresiones < 19829001 | enfermedad pulmonar | : 116676008 | morfología asociada | = << 79654002 |edema|.

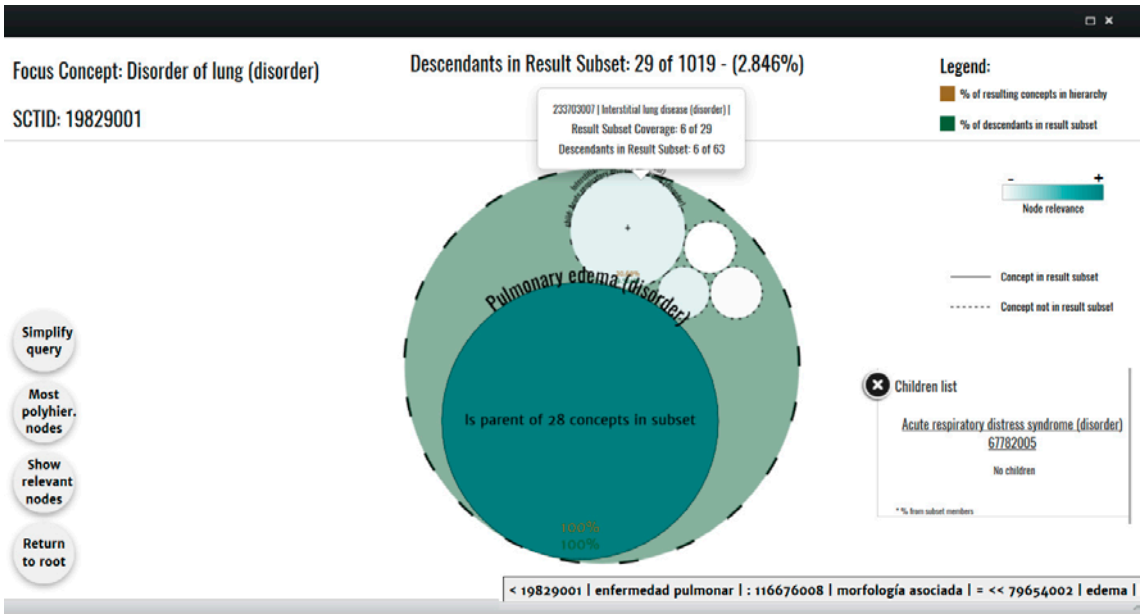

Figura 47. Representación gráfica para el subconjunto de las enfermedades pulmonares asociadas a edema

La tercera representación gráfica es una variante de la anterior, en el sentido que aprovecha la mayoría de la información calculada para llevar a cabo el dibujado. En este caso, en lugar una representación en burbujas, se trata de una representación en forma de árbol. El árbol se representa por niveles horizontales que avanzan en profundidad, de arriba abajo. Asimismo es posible realizar una simplificación de la restricción de expresiones post-ejecución, es posible mostrar n niveles y es posible resaltar los nodos más relevantes. En la figura siguiente se muestra la representación gráfica en forma de árbol de n niveles del subconjunto de las enfermedades pulmonares asociadas a edema.

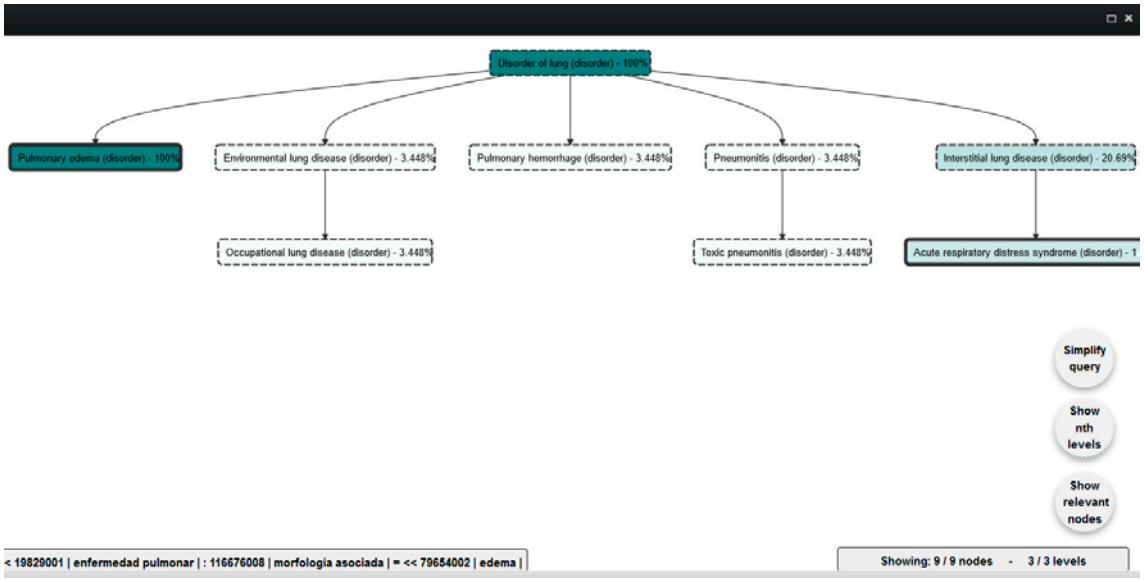

Figura 48. Representación en árbol de n niveles del subconjunto definido por la restricción de expresiones < 19829001 | enfermedad pulmonar | : 116676008 |morfología asociada| = <<79654002 | edema |

9.2.8 Otras funcionalidades

Histórico

La interfaz web del motor de ejecución proporciona un histórico de las últimas diez restricciones de expresión ejecutadas. El histórico también permite volver a volcarlas a la interfaz para su ejecución automáticamente (por ejemplo, para comparar los resultados obtenidos de diferentes sustratos) o para servir de base o plantilla para definir nuevas restricciones de expresiones a partir de ellas.

Plurilingüe

El motor soporta tres idiomas en este momento: inglés, español y valenciano. Debe tenerse en cuenta que este selector de lenguaje afecta a la interfaz, no al sustrato SNOMED CT.

Múltiples sustratos

En este momento, el motor de ejecución admite varias ediciones oficiales de RF2 de SNOMED CT: edición internacional, edición en español y edición del Sistema Nacional de Salud español (SNS). Además, soporta varias versiones de cada edición (es importante recordar que SNOMED International lanza una nueva versión de SNOMED CT cada seis meses).

Conversión entre sintaxis corta y larga

El lenguaje de restricciones de expresiones de SNOMED CT ofrece dos sintaxis lógicamente equivalentes. Por un lado, se considera que la sintaxis corta es la sintaxis normativa y su finalidad es ser lo más compacta posible. Por otro lado, la sintaxis completa o larga utiliza el inglés como alternativa a los símbolos que se definen en la sintaxis corta. El motor de ejecución proporciona un mecanismo para convertir de sintaxis corta a sintaxis larga y viceversa. Debe tenerse en cuenta que los conceptos y atributos de las restricciones de expresiones se completan con sus descripciones completas cuando se utiliza este proceso de conversión.

9.2.9 Tiempos medios de ejecución

Se ha llevado a cabo un análisis de los tiempos medios de ejecución del motor de ejecución en un servidor Windows Server 2008 R2 Datacenter, Intel Core i7-2600K 3,40 GHz y 16 GB de RAM. Específicamente, cada tiempo de ejecución se ha calculado como el promedio de ejecutar diez veces cada consulta. Cabe señalar que el motor utiliza un enfoque de carga perezosa (o “lazy”) incluso para el caso "*". La conclusión más obvia es que el tiempo de ejecución no solo depende del número de conceptos recuperados, sino también de la complejidad de la restricción de expresión y del tamaño de las jerarquías implicadas.

Hay que tener en cuenta que para el tiempo de ejecución de *Id + descripción* (última fila de la tabla), las consultas "*" y C1 se han evaluado sólo en una fase. Por el contrario, las consultas C2-C6 se han evaluado en dos fases para el Id y la descripción, respectivamente.

Las siguientes tablas muestran las seis restricciones de expresiones que se han utilizado para el cálculo de los tiempos medios de ejecución y los tiempos de ejecución (en milisegundos) de las seis restricciones, más una consulta contra todo el sustrato SNOMED CT (es decir, "*"), respectivamente.

	Descripción	Restricciones de expresiones (consultas)
C1	Descendientes de hallazgos clínicos	*< 404684003 |Clinical finding|*
C2	Hallazgos clínicos en la válvula pulmonar	*< 404684003 |Clinical finding| : 363698007 |Finding site| = << 39057004 |Pulmonary valve structure|*
C3	Hallazgos clínicos en la válvula pulmonar asociados a una estenosis	*< 404684003 |Clinical finding| : 363698007 |Finding site| = <<39057004 |Pulmonary valve structure|, 116676008 |Associated morphology| = << 415582006 |Stenosis|*
C4	El mismo subconjunto que 3 pero con dos o más localizaciones en la válvula pulmonar	*<404684003 |Clinical finding|: [2..*] 363698007 |Finding site|=<<39057004 |Pulmonary valve structure|, 116676008 |Associated morphology| = << 415582006 |Stenosis|*
C5	Hallazgos clínicos en la válvula pulmonar asociados con una morfología distinta a estenosis	*< 404684003 |Clinical finding| : 363698007 |Finding site| = << 39057004 |Pulmonary valve structure|, 116676008 |Associated morphology| != << 415582006 |Stenosis|*
C6	Sustancias causantes de hallazgos clínicos	*<<105590001|Substance|: R 246075003 |Causative agent|=<<404684003|Clinical finding|*

Tabla 8. Las seis restricciones de expresiones de SNOMED CT que se han ejecutado en el motor de ejecución para el cálculo de los tiempos medios de ejecución

Consultas (ver tabla 8)	**"*"**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**
Resultados (número de conceptos)	*319446*	*103911*	*116*	*22*	*2*	*88*	*2392*
***Tiempo de evaluación del Id* (ms)**	8811	4212	854	1592	2039	12325	4999
***Tiempo de evaluación de Id + descripción* (ms)**	9266	4744	948	1639	2076	12390	5489

Tabla 9. Tiempos medios de ejecución de las seis consultas (en milisegundos)

10. Conclusiones y trabajo futuro

El motor de ejecución descrito en este trabajo ha sido presentado en varios congresos de investigación, incluyendo el Medical Informatics Europe (MIE 2016). Asimismo, el motor está incluido en la página de implementaciones del lenguaje de restricciones de expresiones de SNOMED CT de la organización que se encarga de mantener la terminología, SNOMED International[9]. También es el motor de ejecución de referencia en los cursos que imparte SNOMED International en su plataforma abierta de e-learning[10]. El motor, asimismo, ha servido de base para el desarrollo de un Trabajo Final de Máster de la Universidad Internacional de la Rioja (UNIR) en el que se expone una metodología de definición de restricciones de expresiones de SNOMED CT a partir de metáforas visuales [17], a la vez que está siendo utilizado por varios países, incluido España, por parte del servicio de semántica del Servicio Nacional de Salud (SNS) del Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI). Asimismo, el motor ha sido utilizado en las clases de prácticas de aula de la asignatura de Sistemas de Información y Telemedicina I y II del Grado en Ingeniería Biomédica de la Universitat Politècnica de València.

El presente Trabajo Final de Grado tiene continuidad en la tesis doctoral que está desarrollando el mismo autor de este trabajo en el Grupo de Informática Biomédica (IBIME) del instituto ITACA de la Universitat Politècnica de València. Concretamente, el desarrollo de un motor de ejecución de restricciones de expresiones de SNOMED CT ha sido el trabajo previo a la labor puramente investigadora, que se centra en el enriquecimiento de modelos de información clínicos, en particular arquetipos, con conocimiento del dominio, para medir la consistencia de las Historias Clínicas Electrónicas. Para enriquecer los arquetipos con conocimiento del dominio es necesario poder definir restricciones de consistencia semánticas y, en particular, enlaces terminológicos de contenido condicional. Tanto para el enlace de contenido como para el enlace de contenido condicional es necesario tener la capacidad de poder crear subconjuntos de conceptos clínicos. Y aquí es donde se pone de manifiesto la necesidad de crear un motor de ejecución de restricciones de expresiones de SNOMED CT, para, seguidamente, poder centrar la tesis en el desarrollo de un lenguaje para expresar conocimiento del dominio de manera formal, y poder así definir las restricciones semánticas en los modelos de información en aras de incrementar su consistencia.

Para la implementación del motor se han estudiado diversas posibilidades en relación al almacenamiento de la base de datos de SNOMED CT. La primera posibilidad ha sido una base de datos relacional y el lenguaje SQL como lenguaje de consulta. No obstante, esta opción se ha desestimado por lo complejo de las consultas al traducir el lenguaje de restricciones al lenguaje SQL y por la poca eficiencia a la hora de consultar los datos. La manera natural de almacenar los conceptos de SNOMED CT y sus relaciones se ha establecido que sea un grafo, dada la estructura de grafo dirigido acíclico de SNOMED CT. Nuevamente se abre el abanico de posibilidades, como crear una base de datos de grafos RDF (Resource Description Framework) y utilizar el SPARQL (Protocol and RDF

[9] https://confluence.ihtsdotools.org/display/SLPG/SNOMED+CT+Expression+Constraint+Language
[10] https://elearning.ihtsdotools.org/course/view.php?id=15#section-1

Query Language) como lenguaje de consultas, o bien, crear una base de datos XML y utilizar el lenguaje XQL (XML Query Language) para las consultas. Finalmente se ha establecido que el lenguaje Cypher del sistema de bases de datos orientadas a grafos Neo4j es susceptible de ser utilizado en la traducción del lenguaje de restricciones de expresiones. Así pues, tras una serie de pruebas, se ha optado por almacenar el sustrato de SNOMED CT en una base de datos de Neo4j y utilizar su lenguaje de consultas, Cypher.

El motor, poco a poco se ha enriquecido con cuestiones de optimización en el almacenamiento (memoria caché), optimización en las consultas, interfaz gráfica, servicio web, opciones de visualización gráfica de los subconjuntos, comparación entre versiones de bases de datos, histórico de consultas, etc. Todas estas tareas se han realizado gracias al apoyo de los investigadores del Grupo IBIME del instituto ITACA de la Universitat Politècnica de València. El siguiente paso es integrar el motor de ejecución en la plataforma de software LinkEHR [18] para el modelado y normalización de HCE basado en arquetipos. EL objetivo es extender esta plataforma con enlaces terminológicos avanzados, dando soporte tanto al enlace semántico como al enlace de contenido entre arquetipos y SNOMED CT.

11. Referencias

[1] Muñoz Carrero A, Romero Gutiérrez A, Marco Cuenca G, Abad Acebedo A, Cáceres Tello J, Sánchez de Madariaga R, Serrano Balazote P, Moner Cano D, Maldonado Segura JA. Manual práctico de interoperabilidad semántica para entornos sanitarios basada en arquetipos. Unidad de investigación en Telemedicina y e-Salud. Instituto de Salud Carlos III. 2013.

[2] openEHR. An open domain-driven platform for developing flexible e-health systems [Internet]. [acceso 19 de agosto de 2017]. openEHR Modelling Tools. Disponible en: http://www.openehr.org/downloads/modellingtools.

[3] openEHR. An open domain-driven platform for developing flexible e-health systems [Internet]. [acceso 10 de agosto de 2017]. Archetype Definition Language 1.4 Specification. Disponible en: http://www.openehr.org/releases/AM/latest/docs/ADL1.4/ADL1.4.html.

[4] SNOMED International. Leading healthcare terminology worldwide [Internet]. [acceso 1 de julio de 2017]. SNOMED CT Starter Guide. Disponible en: https://confluence.ihtsdotools.org/display/DOC

[5] SNOMED International. Leading healthcare terminology worldwide [Internet]. [acceso 1 de julio de 2017]. SNOMED CT Compositional Grammar Specification and Guide. Disponible en: https://confluence.ihtsdotools.org/display/DOC

[6] SNOMED International. Leading healthcare terminology worldwide [Internet]. [acceso 1 de julio de 2017]. SNOMED CT Expression Constraint Language Specification and Guide. Disponible en: https://confluence.ihtsdotools.org/display/DOC

[7] Berges I, Bermudez J, Illarramendi A. Binding SNOMED CT Terms to Archetype Elements. Methods of Information in Medicine. 2014;54(1):45–9.

[8] García MM, Allones JLI, Hernández DM, Iglesias MJT. Semantic similarity-based alignment between clinical archetypes and SNOMED CT: An application to observations. International Journal of Medical Informatics. 2012;81(8):566–78.

[9] Qamar R, Rector A. Semantic Mapping of Clinical Model Data to Biomedical Terminologies to Facilitate Interoperability. U of Manchester. Print. 2008.

[10] Yu S, Berry D, Bisbal J. An Investigation of Semantic Links to Archetypes in an External Clinical Terminology through the Construction of Terminological "Shadows". IADIS Freiburg, Germany. 2010.

[11] Campbell WS, Pedersen J, Mcclay JC, Rao P, Bastola D, Campbell JR. An alternative database approach for management of SNOMED CT and improved patient data queries. Journal of Biomedical Informatics. 2015;57:350–7.

[12] Neo4j, the world's leading graph database - Neo4j Graph Database [Internet]. [acceso 15 de julio de 2017]. Graph Academy Learn Graph Deploy. Disponible en: https://neo4j.com/graphacademy/

[13] Neo4j, the world's leading graph database - Neo4j Graph Database [Internet]. [acceso 23 de julio de 2017]. Intro to Cypher. Disponible en: https://neo4j.com/developer/cypher-query-language/

[14] Robinson I, Webber J, Eifrem E. Graph Databases: new opportunities for connected data. O'Reilly. Second Ed. 2015.

[15] Giménez-Solano VM, Maldonado Segura JA, Salas-García S, Boscá D, Robles M. Implementation of an execution engine for SNOMED CT Expression Constraint Language. Pages 466 - 470. DOI:10.3233/978-1-61499-678-1-466. Series Studies in Health Technology and Informatics. Ebook Volume 228: Exploring Complexity in Health: An Interdisciplinary Systems Approach.

[16] JavaCC. The Java Parser Generator [Internet]. [acceso 10 de julio de 2017]. JavaCC Documentation. Disponible en: https://javacc.org/doc

[17] De la Asunción González E. Development of Web Applications to edit SNOMED-CT expressions [tesina de máster]. Universidad Internacional de La Rioja. 2016.

[18] Maldonado JA, Moner D, Boscá D, Fernández-Breis JT, Angulo C, Robles M. LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. International Journal of Medical Informatics. 2009;78(8):559–70.